Bad Boys: How Criminal Identity Affects Rule Violation*

Alain Cohn, Michel André Maréchal and Thomas Noll

November, 2013

Abstract

We conducted an experiment with 182 inmates from a maximum security prison to analyze the impact of criminal identity on cheating. The results demonstrate that inmates cheat more when we render their criminal identity more salient. This effect is specific to individuals possessing a criminal identity, because an additional placebo experiment shows that regular citizens do not become more dishonest in response to crime related reminders. Moreover, our experimental measure of cheating correlates with inmates' offenses against in-prison regulation. Altogether, these findings suggest that criminal identity plays a crucial role in rule violating behavior.

JEL classification: K00, C93, K14, K42, Z10
Keywords: Dishonesty, Identity, Crime, Prison, Experiment.

* Cohn: University of Zurich, Department of Economics, Bluemlisalpstrasse 10, CH-8006 Zurich, Switzerland, email: alain.cohn(at)econ.uzh.ch. Maréchal: University of Zurich, Department of Economics, Bluemlisalpstrasse 10, CH-8006 Zurich, Switzerland, email: michel.marechal(at)econ.uzh.ch. Noll: Swiss Prison Staff Training Center, Avenue Beauregard 11, CH-1700 Freiburg, Switzerland, email: noll(at)prison.ch. We thank Milena Brunner, Marion Cohn, Markus Hauzenberger, Bigna Ruppen, Giuseppe Ugazio and Sophia Zügel for providing excellent research assistance and David Huffman for inspiring discussions. We are grateful to Henk Aarts, John Antonakis, Dan Ariely, Daniel Benjamin, Sam Bowles, Alexander Cappelen, Daniel Chen, Stefano DellaVigna, Jérôme Endrass, Ernst Fehr, Herb Gintis, Sally Gschwend, Dan Kahan, Martin Killias, Ilyana Kuziemko, Richard McAdams, Nick Netzer, Jesse Shapiro, Bertil Tungodden, Glenn Walters and Roberto Weber, as well as the audiences at Invitational Choice Symposium Erasmus, SITE Stanford, ESA Chicago, University of Chicago Booth, Deception, Incentives & Behavior Symposium San Diego, MPI Bonn, WZB Berlin, THEEM Konstanz, University/ETH Zurich, and Forensic Psychology & Psychiatry Canton Zurich for helpful comments. We acknowledge the support of the Gottlieb Duttweiler Institute. Correspondence to: Cohn and Maréchal.

1 Introduction

Prison populations have skyrocketed in most parts of the world over the past decades. Worldwide, more than 10.1 million people are currently kept in penal institutions (Wamsley 2011). The fact that prison doors revolve for many inmates exacerbates the problem of prison overcrowding. Estimates from the United States, for example, suggest that more than four out of ten inmates who are released are rearrested within three years (Pew Center on the States 2011).

Despite a longstanding scientific discussion, the consequences of imprisonment on subsequent delinquency are controversial. According to the economic theory of crime, pioneered by Becker (1968), delinquent behavior depends on its expected benefits and costs. To the extent that delinquents tend to underestimate the costs or likelihood of punishment (Lochner 2007), the aversive experience of imprisonment should deter ex-prisoners from re-offending (see Smith and Gartin 1989).[1] Other theories from economics and sociology suggest that incarceration ingrains a criminal identity which is assumed to promote delinquency. Proponents of the prisonization theory, for example, argue that imprisonment places individuals in a unique social environment - the so called "society of captives" (Sykes 1958) - which has its own informal inmate code often consisting of values and norms antithetical to the society outside of prison. Exposure to this subculture can lead to the internalization of "the folkways, mores, customs and general culture of the penitentiary" (Clemmer 1940, p. 299) and strengthens criminal identity.[2] Lerman (2009), for example provides quasi-experimental evidence for the influence of imprisonment on criminal identification. Exploiting discontinuities in the assignment to security levels, she finds that harsher prison conditions promote criminal personality in inmates. Moreover, Walters (2003) finds that the reinforcement process is quite fast: most of the iden-

[1] This line of reasoning is known as the *specific* deterrence hypothesis. By contrast, the hypothesis of *general* deterrence postulates that the threat of imprisonment deters people from criminal activities in the first place (see Levitt 2002). Note that the two mechanisms are not mutually exclusive.

[2] Labelling theorists argue in similar spirit that a person's identity can change if society treats him or her as a criminal, leading to the adoption of behavioral propensities consistent with the criminal label. For example, Tannenbaum claimed that ultimately "the person becomes the thing he is described as being" (1957, p. 20).

tity change appears in the first months of imprisonment.[3] While there is evidence that imprisonment fosters inmates' criminal identity, the critical but still unanswered question is whether criminal identity actually influences rule violating behavior.[4] We fill this gap in empirical knowledge and provide causal evidence for the effect of criminal identity on cheating, as a form of rule-violating behavior.

Identifying the causal influence of criminal identity is challenging. A simple comparison of behavior in criminals and non-criminals can be misleading due to omitted variables.[5] Criminals and law-abiding citizens differ in many dimensions which are often unobservable to researchers, such as their financial background, life prospects, or opportunity costs of time. We therefore opted for a different approach based on the economic theory of identity developed by Akerlof and Kranton (2000) which posits that people have multiple social identities such as their gender, ethnicity, or social class. Identities are tied to different norms which prescribe how people should behave. According to Akerlof and Kranton, identity influences economic choice because deviating from the identities' prescribed behavior causes disutility. Which identity norms dominate in a given situation depends on the relative weight (i.e. salience) an individual attaches to a specific identity (see Turner 1985). In our context, we assume that prisoners have a criminal identity and a moral identity. Dishonesty inflicts psychological costs on the moral identity (e.g. see Gneezy 2005, Charness and Dufwenberg 2006, or Mazar et al. 2008), but not on the criminal identity. Accordingly, if the criminal identity receives more weight, inmates should become more dishonest.

We test whether criminal identity causes cheating by conducting an experiment with 182 inmates from the maximum security prison Pöschwies - Switzerland's largest penitentiary for male adults. We randomly increased

[3]Ethnographic studies document that prisoners even speak their own language, "prison argot" (see Clemmer 1940, Kaminski 2004), which is a further indication for the development of a common social identity (Gumperz 1982).

[4]Several sociological studies document correlations between survey measures of criminal identity and deviant behavior (see Gendreau et al. (1996) for a meta-analysis.). These studies, however, fail to show whether criminal attitudes influence crime or whether causality runs the other way round.

[5]See Birkeland et al. forthcoming, Khadjavi and Lange 2013, and Chmura et al. 2013 for studies comparing distributive preferences and cooperation in criminals and non-criminals.

the saliency of criminal identity in half of the participants using embedded survey questions that reminded them of the fact that they are incarcerated criminals (e.g. "What were you convicted for?"). The other half of the participants served as the control group and answered questions unrelated to their criminal identity (e.g. "How many hours per week do you watch television on average?"). Subsequently, we measured inmates' dishonesty in a simple coin tossing task. The rules of the task required subjects to flip ten coins and report the outcome on paper. They were allowed to keep every coin for which they reported "heads", creating a monetary incentive to break the rules and misreport the coin flips. Because participants were unobserved they could easily hide behind chance and thus did not have to fear any punishment. We are, nevertheless, able to measure cheating at the group level, as we know the distribution which should result from honest reporting.[6]

The results show that many of the inmates cheated. On average, they reported heads for 60 percent of the coin flips in the control condition. This is significantly above chance and approximates 20 percent of misreported coin flips. Inmates became even more dishonest when we rendered their criminal identity more salient. In the criminal identity treatment they reported 66 percent of heads, which corresponds to 32 percent of misreported coin flips. Thus, the higher saliency of criminal identity increased the frequency of misreporting by 60 percent. Using administrative data, we further show that behavior in the coin tossing task correlates with inmates' offenses against in-prison regulation (e.g. aggression against others, use of illegal drugs, or weapon possession), suggesting that the coin tossing task provides a valid measure of rule violating behavior. Half a year after the main study, we conducted a further experiment in the same prison, which serves as a manipulation check. Based on an implicit measure of criminal cognition we find that the criminal identity questions enhanced the mental accessibility of crime-related thoughts. This indicates that the treatment manipulation worked as intended.

We discard several alternative explanations to a criminal identity effect. For example, one could argue that the questions might have triggered

[6]See Fischbacher and Föllmi-Heusi (2013), or Abeler et al. (2012) for similar approaches for eliciting dishonest behavior.

negative emotions and arousal because they reminded prisoners about their criminal deeds or the social injustice of incarceration (Sherman 1993). Such an emotional reaction could potentially undermine their honesty. However, we demonstrate that the criminal identity questions did not influence emotions, and that the correlation between emotions and cheating is insignificant in any case. Furthermore, the saliency of criminal identity might have influenced inmates' risk attitudes (Benjamin et al. 2010). If inmates erroneously believed that cheating was individually detectable, a decrease in risk aversion could also explain a higher cheating rate in the criminal identity treatment. We therefore elicited inmates' risk attitudes and tested whether they are correlated with earnings in the coin tossing task. The correlation is close to zero and statistically insignificant, suggesting that a change in risk attitudes is unlikely to drive the treatment effect.

Finally, we conducted an additional "placebo" experiment with regular citizens in order to further consolidate the criminal identity effect. A non-criminal population should not respond to crime-related questions, because this group does not possess a criminal identity. We tested this prediction with 193 male citizens recruited from the general population. The survey administered to the general population was identical to the prisoner survey, except for the treatment manipulation, which needed to be adapted slightly. Before answering the six crime-related or control questions, participants memorized a short text profile describing either a criminal or a non-criminal person. We created these profiles using representative answers from the prisoners in the criminal identity, respectively the control treatment. Subsequently, subjects answered the same questions as the prisoners from the perspective of the person described in the profile. The results show that the criminal profile treatment had no significant influence on cheating. If anything, the effect goes in the opposite direction: the fraction of misreported coin flips drops from 14 percent in the control condition to 10 percent in the criminal condition. Altogether, the three studies suggest that criminal identity promotes rule violating behavior.

Our results contribute to a growing literature studying the role of social identities in economic decision making (e.g. see Akerlof and Kranton 2000, 2008, Fang and Loury 2005 or Bénabou and Tirole 2011). The empirical literature mostly analyzed whether people discriminate whether their

interaction partners belong to the in-group and share the same social identity (e.g. Hoff and Pandey 2006, Charness et al. 2007, Chen and Li 2009, Goette et al. 2012, or Kranton et al. 2012). One of the few exceptions is the study by Benjamin et al. (2010) who used a similar approach to ours and analyzed the influence of ethnic and gender identity salience on risk and time preferences. A more recent study by Bertrand et al. (2013) illustrates that gender identity norms influence a wide range of economic and social outcomes. They find that the norm that "the wife should not earn more than the husband" helps explain labor force participation and relative income, divorce and the division of home production within US households.[7] We add to this literature by providing first causal evidence for the impact of social identity on dishonest behavior. In this sense our results are also relevant for the rapidly expanding literature on the determinants of dishonesty (see Gneezy 2005, Charness and Dufwenberg 2006, Mazar et al. 2008, Shalvi et al. 2011, Fischbacher and Föllmi-Heusi 2013, or Pruckner and Sausgruber 2013).

Our study further speaks to a large literature studying the effects of imprisonment and prison conditions on recidivism. Overall the evidence points towards a null effect or even a positive effect of imprisonment on in recidivism (see Nagin et al. 2009 for a systematic review). However, this evidence needs to be taken with a grain of salt: imprisonment is selectively imposed and failing to account for unobserved heterogeneity between convicts under different sanctioning regimes can be misleading (Manski and Nagin 1998). Few studies allow to draw causal inference by using experimental or quasi-experimental data. Chen and Shapiro (2007), for example, identify the effect of prison conditions on recidivism based on discontinuities in the assignment of inmates to prison security levels. Their regression discontinuity analysis suggests that if anything harsher prison conditions tend to increase recidivism. Another study by Killias et al. (2000) analyzes a randomized field experiment comparing the impact of short term imprisonment and community service. They found that short term imprisonment increases recidivism.[8] Several explanations have been put forth as to why

[7] See also Alesina et al. (2013) for a recent study on the historical origins of today's gender identity norms.

[8] Aizer and Doyle (2013) and Drago et al. (2011) provide consonant evidence. By contrast, the results from Kuziemko (2013) and Hjalmarsson (2009) suggest that impris-

imprisonment might increase recidivism. Ex-prisoners may, for example, face difficulties when re-integrating into society due to the social stigma of imprisonment (Western et al. 2001). They are often discriminated in important aspects of their social and professional lives, such as in their job search (Pager 2003, Falk et al. 2009). Moreover, prisons are commonly viewed as "schools for crime" where inmates learn new crime methods and opportunities or expand their criminal networks (Bayer et al. 2009). Our findings complement this literature by providing an additional mechanism through which imprisonment could increase criminal activity. At this point we want to emphasize that our results should not be taken as evidence against the effectiveness of prisons in general. Prisons incapacitate dangerous and habitual criminals (see Kessler and Levitt 1999, Buonanno and Raphael forthcoming) and the mere threat of imprisonment can deter people from committing crime (see Levitt 1998, Helland and Tabarrok 2007, Lee and McCrary 2009, or Drago et al. 2009). However, we believe that criminal identity is an important aspect for policy makers and practitioners to consider in the organization of everyday life in prison as well as in designing therapeutic programs.

Finally, our approach is conceptually related to the economic literature on salience and limited attention (e.g. Chetty et al. 2009, Bushong et al. 2010, Lacetera et al. 2012, Gabaix 2013, or Bordalo et al. forthcoming), which has in common that the agent's focus affects the decision weight given to different attributes. While this literature is concerned about the weights given to attributes in the choice set, in our study we manipulate the weight of a specific social identity.

2 An Experiment Behind Bars

Design

We conducted our experiment in the maximum security prison Pöschwies - Switzerland's largest penitentiary for male adults. A total of 182 inmates participated in the experiment. The majority of them were convicted of violent crimes (30%), followed by drug related crimes (26%), property crimes

onment decreases recidivism.

(24%), sex crimes (15%), and other types of crime (5%). Almost two third (62%) were repeat offenders. Participants had been incarcerated for 2.7 years on average, with a minimum of 26 days and a maximum of 22.5 years.[9]

We sent an invitation for a survey study from the University of Zurich to all inmates. Participants were assured of confidentiality and that their individual data would not be revealed to the prison authorities. Interested participants could choose their preferred survey language among four options: German, English, Italian, and French. A few days later, participants received an envelope containing the survey (see online appendix), and a second smaller envelope they were instructed to open at a later point in time. We ensured that the inmates completed the survey in private without being disturbed by guards or other inmates. Inmates from single cells received the survey overnight, and those who shared their cell with another inmate completed the survey while their cellmate was working.[10] The experiment was conducted over one period of 24 hours in order to minimize the possibility of talking about it.

The first part of the survey contained filler questions about subjective wellbeing and standard demographics. The second part comprised our key experimental manipulation. We randomly primed half of the participants with their criminal identity by asking them six questions that reminded them of the fact that they are incarcerated criminals (e.g. "What were you convicted for?" or "How long have you been in custody?").[11] The other half of the participants served as the control group and answered six questions unrelated to their criminal identity (e.g. "What is your favorite activity when you do not have to work?" or "How many hours per week do you watch television on average?"). These six questions were the only

[9]See Table A1 in the appendix for further descriptive statistics of our sample. Table A4 in the appendix shows descriptive statistics from the whole prison population in the year of the experiment. The composition of participants is very similar to that of the total prison population.

[10]Working hours are staggered for inmates in double cells. Each inmate works for half a day.

[11]Priming is a method developed in psychology and refers to the activation of mental representations through situational cues (Bargh and Chartrand 2000, Shih et al. 1999, or LeBoeuf et al. 2010). Priming is now increasingly used in economics (e.g. Callen et al. forthcoming, Hoff and Pandey forthcoming, Benjamin et al. 2010, 2013, or Chen et al. 2010).

difference between the criminal identity and the control treatment.[12] Immediately after the priming, participants were asked to indicate their current emotional state using non-verbal Self-Assessment Manikins (Bradley and Lang (1994)).[13] This allows us to identify potential emotional reactions to the priming questions.

Towards the end of the survey, subjects were instructed to open the second envelope, which contained ten coins, each worth 0.5 Swiss francs (or 0.55 US dollars). The rules required the participants to flip the coins sequentially and to report the outcome on paper. They were allowed to keep every coin for which they reported "heads". If they flipped tails, they had to put the coin back into the envelope together with the survey, which had to be handed over to the guards on the next morning. Participants thus had a monetary incentive to cheat by misreporting the outcome of their coin flips. The stake size was sizable for the participants, considering that the maximum payoff matched their hourly wage in prison. Because participants could hide behind chance, it is impossible to determine with certainty whether an individual cheated or not. They therefore did not have to fear any adverse consequences from cheating. However, we are able to infer the extent to which participants in different groups cheated by comparing the empirical distributions of reported heads with the binomial distribution implied by honest behavior. Moreover, assuming that none of the participants cheated to his disadvantage - reporting tails when the actual outcome was heads - we are able to calculate the percentage of misreported coin flips (see also Houser et al. 2012). Let h be the percentage of reported heads and m the percentage of misreported coin tosses. The percentage of reported heads is therefore determined by

$$h = m * 1 + (1 - m) * 0.5 = 0.5 * (1 + m). \tag{1}$$

If a participant cheats, he reports heads with probability 1. However, honest reporting implies that heads occurs only with probability 0.5. We can thus characterize the percentage of misreported coin tosses by

[12] In order to ensure a ceteris paribus comparison of the two treatments, we also matched the answer formats of the two sets of questions.

[13] These measures have been shown to be consistently correlated with different physiological measures, such as heart rate and facial muscle contraction (Bradley and Lang 2000).

$$m = 2 * h - 1. \tag{2}$$

The last column of Table A1 in the appendix reports whether there are any systematic differences between participants in the two treatment conditions and serves as a randomization check. The background characteristics appear well-balanced across treatments. There are no significant differences between groups in length of sentences, disciplinary offenses, re-offending status, age, cognitive skills[14], risk attitudes or assignment to prison section. Only the fraction of inmates in the conviction category "Other" is significantly lower ($p<0.05$, χ^2-test) in the criminal identity treatment. Treatment differences in the number of years in prison ($p=0.100$, rank-sum test) and the share of inmates who completed compulsory school ($p=0.095$, χ^2-test) are marginally significant. These marginal differences do not occur more frequently than chance would dictate. We nevertheless control for all three variables in our regression analysis.

Framework and Hypothesis

We develop a simple framework, based on Benjamin et al. (2010), to derive our hypothesis. We assume that inmates have a criminal identity and a moral identity. These social identities are tied to norms that prescribe different behavior in the coin tossing task. As a consequence, an internal conflict arises over appropriate behavior. Let $x_i \in [0, 10]$ be the number of heads inmate i reports, and let x_c denote the action his criminal identity prescribes, and x_m the action his moral identity calls for. Following the economic model of crime (Becker 1968), the action prescribed by the criminal identity x_c is derived from maximizing the following utility function:

$$\max_{x_i \in [0,1]} EU_i(x_i) = (1-p)x_i - pf, \tag{3}$$

where p is the detection probability and f the fine imposed on individual i when caught cheating. Given that the probability of detection in our context is zero, i.e., $p = 0$, the utility of the criminal identity is maximized by reporting ten times heads ($x_c = 10$). In contrast to the criminal identity

[14]Cognitive skills were elicited using Frederick's (2005) cognitive reflection test.

that prescribes cheating to the full extent, we assume that cheating imposes psychological costs on the moral identity (e.g. see Ellingsen and Johannesson 2004, Gneezy 2005, Charness and Dufwenberg 2006, and Mazar et al. 2008).[15] Let x_0 denote the true number of tossed heads and let λ denote the psychological cost of cheating. Consequently, the action prescribed by the moral identity x_m is derived from maximizing the following utility function:

$$\max_{x_i \in [0,1]} U_i(x_i, x_0) = x_i - \lambda(x_i - x_0)^2. \tag{4}$$

Solving the maximization problem yields the moral identity's prescribed action $x_m = x_0 + \frac{1}{2\lambda}$, which depends on both the true outcome x_0 and the psychological costs of cheating λ. The larger these psychological costs, the more x_m corresponds to the truth x_0. In the identity model, we assume that a criminal person maximizes a utility function that is a convex combination of the prescribed actions of his criminal and moral identity. Deviating from the prescribed actions causes disutility. A criminal's maximization problem can thus be characterized as follows:

$$\max_{x_i \in [0,1]} U_i = -w(s)(x_i - x_c)^2 - (1 - w(s))(x_i - x_m)^2. \tag{5}$$

where $0 \leq w(s) \leq 1$ is the weight placed on the criminal person's criminal identity, and $1 - w(s)$ is the relative importance of his moral identity. Without loss of generality, we assume that $w(0) = 0$ and $w' > 0$. s is the strength of the criminal identity and has a steady-state value of $\bar{s}$. Environmental cues or primes can temporarily disturb the steady-state by $\epsilon > 0$, i.e., $s = \bar{s} \pm \epsilon$. Inserting the preferred actions x_c and x_m into equation (5) and solving the maximization problem gives the following optimal action for individual i:

$$x_i^* = w(s) \cdot 10 + (1 - w(s))(x_0 + \frac{1}{2\lambda}). \tag{6}$$

The optimal action of a criminal with identity considerations is the weighted average of the prescribed actions of his criminal and moral identity. By temporarily increasing the saliency of criminal identity in our

[15] There are several ways to model lying costs. Our main prediction, however, does not depend on the exact form of the lying costs.

experiment, we augment the strength s of the criminal identity by ϵ. According to equation (6), the optimal action therefore shifts towards the action the criminal identity prescribes. This leads to the following hypothesis:

Hypothesis: *Inmates report, on average, more successful coin flips in the criminal identity treatment than in the control treatment.*

3 Experimental Results

We outline our results in four steps. First, we examine the impact of criminal identity on cheating. Second, we validate our experimental measure of cheating by showing that behavior in the coin tossing task is correlated with inmates' offenses against in-prison regulation. Third, we analyze data from an additional experiment with prisoners, which provides a manipulation check for the identity priming. Finally, we present the result from a placebo experiment conducted with subjects from the general population and test whether the identity effect is specific to criminals.

Criminal Identity and Cheating

The results from the coin tossing task show that many of the inmates cheated. On average, they reported heads for 63 percent of the coin flips, which is significantly above chance (95% confidence interval: [60%, 66%][16]). Assuming that none of the prisoners cheated to his disadvantage - reporting tails when the actual outcome was heads - we estimate that 26 percent of the coin flips were misreported (see Equation 2).

Panel A of Figure 1 shows the binomial distribution of the number of heads which should theoretically result if everyone was honest, and the empirical distribution from the control treatment. The latter is clearly skewed towards a higher number of heads than honest behavior predicts. For example, while we should theoretically expect around 0.1 percent of the participants to win the maximum amount, almost 13 percent of the prisoners

[16]The confidence interval is based on individual averages to account for the fact that reporting behavior could be correlated within individuals.

reported so ($p < 0.001$, Binomial test). The distribution from the criminal identity treatment is even further shifted towards higher payoffs (see panel B). The outcome 10, but also outcomes of 8, 7 and 6 times heads, were significantly more frequent than honest reporting would predict ($p < 0.001$, 0.007, 0.059, and 0.097, Binomial tests). As a result, the average percentage of heads increased significantly from 60 percent in the control group to 66 percent in the criminal identity group ($p = 0.017$, rank-sum test), as shown in panel C.[17] The corresponding rates of misreported coin flips are 20 percent, and 32 percent respectively, suggesting that cheating is 60 percent more frequent in the criminal identity than in the control treatment. Interestingly, most of the treatment effect comes from incomplete cheaters (i.e. those who report 6, 7 and 8 times heads), arguably those who face a stronger tension between their moral and criminal identity.

The regression results in Table 1 are in line with the preceding nonparametric analysis. We estimate a Probit model of the following form:

$$Pr(heads_{ik} = 1) = \Phi(\alpha + \beta * C_i + \gamma * X_i + \epsilon_{ik}). \tag{7}$$

The decision of individual i to report heads for coin toss k is regressed on the criminal identity treatment dummy C_i. We additionally control for the residual category of convictions, the number of years in prison, and compulsory schooling level in the vector X_i, because these variables were imperfectly balanced across treatments. $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

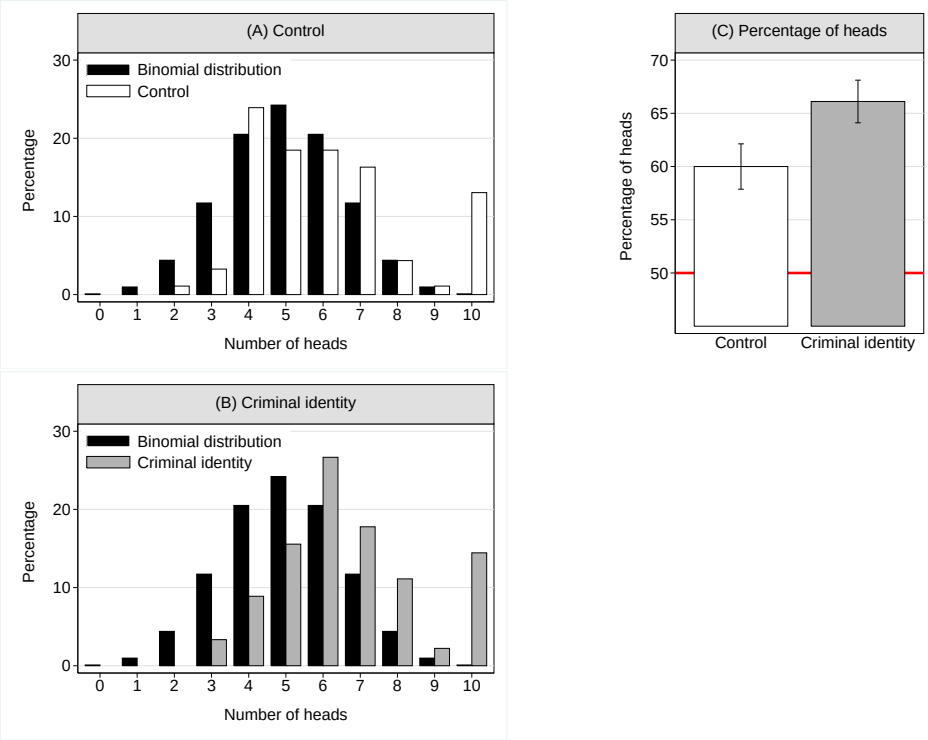
The marginal effect reported in column (1) of Table 1 reveals that the probability of reporting heads in the criminal identity treatment is 6.1 percentage points higher than in the control condition ($p = 0.036$). As shown in column (2) this result remains robust if we include the additional background characteristics as control variables.[18] Our findings are summarized in the following result:

Result 1: *Prisoners cheat more when their criminal identity is made more salient.*

[17] All p-values reported in this paper are based on two-sided tests except for the Binomial tests which are based on directed hypotheses.

[18] We alternatively estimated a linear probability model using Ordinary Least Squares (OLS), yielding the same results.

Figure 1: Criminal identity and cheating in inmates



Panel A of this figure shows the distribution of heads reported by the prisoners in the control treatment and the binomial distribution honest reporting implies. Panel B depicts the distribution of heads in the criminal identity treatment and the binomial distribution. Panel C compares the average percentage of heads reported in the control and in the criminal identity treatments.

We tested the relevance of several alternative interpretations. First, being reminded of one's criminal activity or the social injustice of being incarcerated (Sherman 1993) might have provoked arousal and negative emotions which may have affected dishonesty. We measured participants' arousal and affective state immediately after the priming questions using validated non-verbal Self-Assessment Manikins (Bradley and Lang 1994). As shown in Figure 2, the saliency of criminal identity neither had an effect on arousal ($p = 0.369$, rank-sum test) nor on negative affect ($p = 0.323$, rank-sum test).

Second, the criminal identity manipulation might have altered criminals' risk attitudes (see Benjamin et al. 2010), and thus possibly also their inclination to cheat. Even though it was impossible to detect whether an

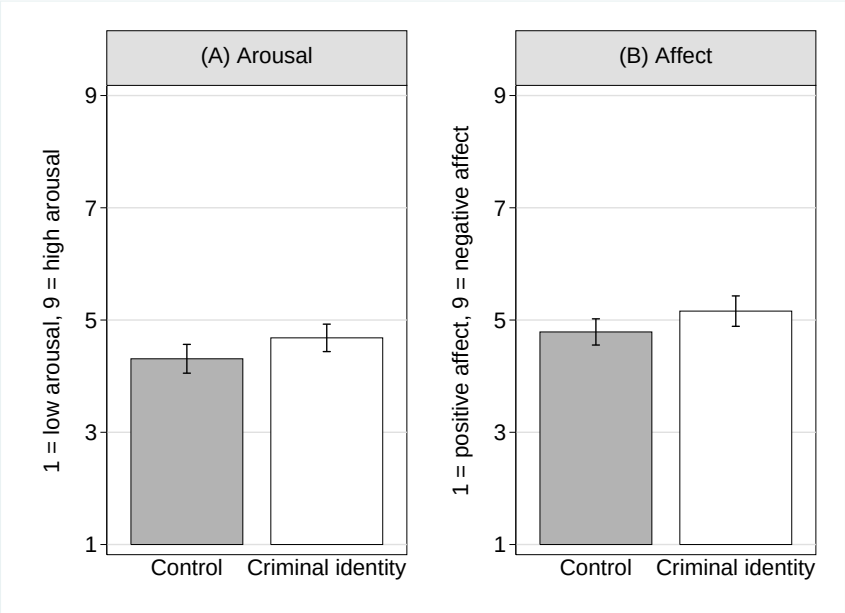
Table 1: Regression analysis: Criminal identity and cheating

Dependent variable:	(1)	(2) heads = 1	(3)
Criminal identity	0.061**	0.068**	0.069**
	(0.029)	(0.030)	(0.030)
Arousal			-0.011
			(0.007)
Negative emotions			0.012
			(0.007)
Risk attitudes			-0.000
			(0.006)
Additional controls:	no	yes	yes
Observations	1820	1730	1630
Subjects	182	173	163

This table reports marginal effects from a probit model, calculated at the median levels of the covariates. Robust standard errors, corrected for clustering on the individual level, are displayed in parentheses. In column (1), the decision to report heads is regressed on the criminal identity treatment dummy. Column (2) additionally controls for the residual category of convictions, the number of years in prison, and compulsory schooling level, because these variables were imperfectly balanced across treatments. Column (3) also controls for arousal, negative emotions, and risk attitudes. Due to item non-response, the number of observations drops when adding covariates. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

individual participant cheated, one could argue that some participants erroneously believed that they might get caught and that this would entail negative consequences for them. We elicited inmates' risk attitudes prior to the priming using an experimentally validated questionnaire measure of risk attitudes (Dohmen et al. 2011). We found no significant relationship between individual risk attitudes and behavior in the coin tossing task (Spearman's $\rho = -0.017$, $p = 0.820$). Furthermore, we re-estimated model (7) and controlled for arousal, negative affect and risk attitudes in column (3) of Table 1. None of the variables reaches statistical significance and the coefficient estimate for the criminal identity treatment remains unchanged. Together, these results suggest that neither emotions nor risk attitudes are able to explain the treatment effect.

Figure 2: Arousal and affect



Panel A (B) of this figure shows the average self-reported arousal (affective state) in the control and criminal identity treatment.

Validity of the Coin Tossing Task

We present complementary evidence showing that the coin tossing task provides a valid measure of rule violating behavior. We were given access to the anonymized administrative records of disciplinary offenses for each participant. Typical offenses are aggression against others, drug or weapon possession, and other kinds of illegal activities. The average inmate had a record of two disciplinary offenses since the beginning of incarceration. We used this information to test whether behavior in the coin tossing task correlates with inmates' institutional behavior and estimated the following model using OLS:

$$y_i = \alpha + \beta * H_i + \gamma * T_i + \delta * X_i + \epsilon_i. \tag{8}$$

We regressed the number of offenses y_i committed by inmate i since incarceration on the percentage of reported heads H_i in the coin tossing task. We control for the different windows of opportunity using the time each criminal spent in prison T_i as an additional explanatory variable. We

further estimated a model, where we control for a large set of additional criminal background measures X_i. The coefficient estimate reported in column (1) of Table 2 suggests that, on average, inmates who reported ten times heads committed two more offenses in prison than those who reported heads in 50 percent of the cases ($p = 0.034$). This difference corresponds to roughly five additional years of imprisonment (see coefficient estimate for "Years in prison"). The regression results are robust to the inclusion of criminal background and socio-economic characteristics, as shown in column (2) and (3).[19] The following result summarizes our findings:

Result 2: *Behavior in the coin tossing experiment correlates with rule violating behavior in prison.*

Manipulation Check

Six months after the main study we conducted a second experiment with 119 inmates from the same prison using the same procedure. The goal of this follow-up experiment was to measure the impact of the priming questions on criminal cognition, which serves as a manipulation check.

The first part of the follow-up survey included new filler questions, mostly on subjective wellbeing (see online appendix). The second part contained exactly the same six priming questions as in the previous experiment.[20] Following the six questions, participants solved a word stem completion task. For example, they could complete the word stem "off..." with the crime-related word "offense" or unrelated words such as "office". The other two word stems were "acc..." (e.g. accusation vs. account) and "pol..." (e.g. police vs. politics). A research assistant who was blind to the experimental conditions categorized the answers into crime-related and

[19] We alternatively treated disciplinary offenses as count data and estimated a negative binomial regression model, yielding similar results.

[20] 69 inmates had already participated in the first experiment; the other 50 inmates participated for the first time. Table A2 in the appendix shows the descriptive statistics of the participants of the manipulation check. The randomization check in the last column suggests that all background characteristics are well balanced across treatments. We reversed treatment assignment for subjects who already participated in the main experiment: those who originally answered the criminal identity questions received the control questions instead, and vice versa. Treatments were randomly assigned for subjects who participated for the first time.

Table 2: Prison rule violations and behavior in the coin tossing task

	(1)	(2)	(3)
Dependent variable:	# of disciplinary offenses		
Percentage of heads	0.040**	0.037**	0.037**
	(0.019)	(0.018)	(0.018)
Years in prison	0.404***	0.279**	0.433***
	(0.109)	(0.109)	(0.140)
Criminal background controls:			
Type of conviction	no	yes	yes
Repeat offender	no	yes	yes
Prison section	no	yes	yes
Socio-economic controls:			
Age	no	no	yes
Nationality	no	no	yes
Education	no	no	yes
Observations	182	182	159

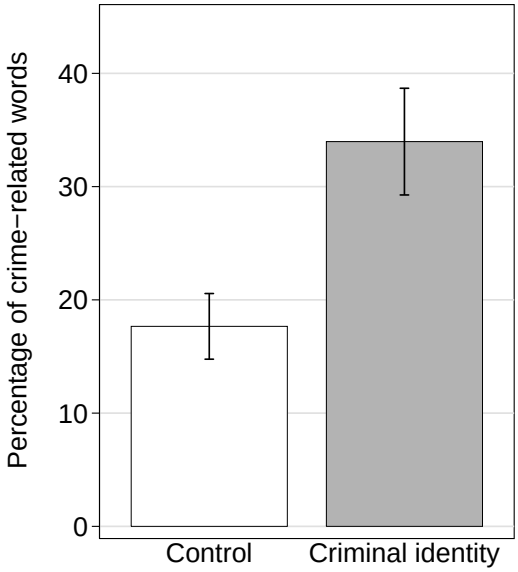
This table reports OLS coefficient estimates. Robust standard errors are displayed in parentheses. In column (1), the number of disciplinary offenses is regressed on the percentage of heads reported and the number of years in prison. Column (2) includes additional criminal background controls, such as type of conviction, repeat offender status, and prison section. We do not control for sentence length because it is not determined for more than one-third of the sample (early imprisonment and safe custody). Column (3) also controls for age, nationality, and education. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

unrelated words. This allows us to compare the mental accessibility of crime-related constructs across treatments.

As depicted in Figure 3, the mental accessibility of crime-related constructs was effectively manipulated. In comparison to the control condition, the participants in the criminal identity treatment mentioned crime-related words almost twice as frequently ($p = 0.008$, rank-sum test).

Result 3: *In comparison with the control treatment, the criminal identity treatment increased the mental saliency of crime-related constructs, suggesting that the treatment manipulation worked as intended.*

Figure 3: Manipulation check



This figure shows the average percentage of crime-related words in the word completion task by treatment condition.

Placebo Experiment with Regular Citizens

We conducted an additional placebo experiment with subjects from the general population in order to further consolidate our results with an effect of criminal identification. A criminal identity effect implies that a non-criminal population should not respond to crime-related questions, because this group does not possess the corresponding identity.

To test this prediction, we recruited 193 male visitors from the resident registration office of a Swiss municipality.[21] We deliberately chose a municipality characterized by a relatively high proportion of foreigners in order to recruit participants with a similar cultural background to that of the prisoners. Moreover, recruitment at the registration office allowed us to approach representative citizens of the chosen community.[22] The

[21] We cannot rule out that some of our subjects might have committed crime because we do not know their criminal histories. However, if some of the subjects are indeed criminals, their presence would work against our prediction, making the test even stronger.

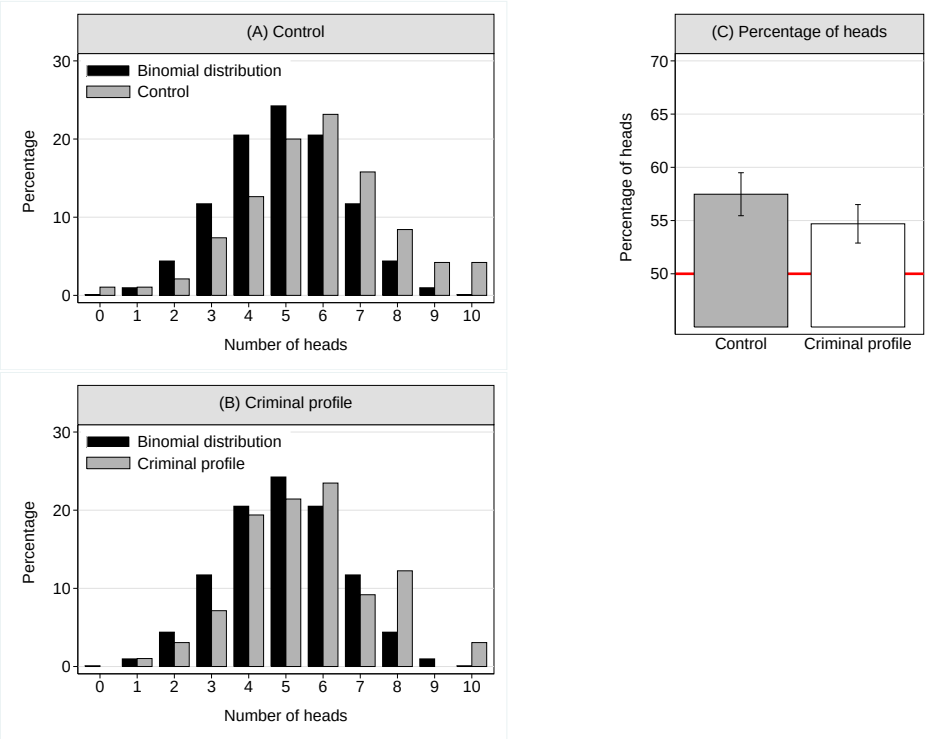
[22] Citizens mostly visit the registration office to receive or renew official documents, such as passports and residency permits.

experimental design and procedure followed the prisoners' experiment as closely as possible. Participants were assured that their answers would be treated confidentially, and they could choose their preferred survey language among the same four options. Subjects received an envelope that contained the survey, and a second smaller envelope with ten 0.5 Swiss franc coins. Participants filled out the survey alone in an empty room at the resident registration office, ensuring the same degree of privacy during the experiment. They received an additional fixed show up fee of 10 Swiss francs in order to compensate them for their higher opportunity costs of time. The survey was identical to that administered to the prisoners, except for the treatment manipulation, which needed to be adjusted slightly. Before answering the crime or control questions, participants had to memorize a short text profile of a person. They were randomly assigned to a profile of a criminal or a non-criminal person. We created these profiles using representative answers from the prisoners in the criminal identity and the control treatment, respectively (see online appendix). Subsequently, subjects answered the same six questions as the prisoners from the perspective of the person described in the profile. After the quiz, participants were asked to open the second envelope and to complete the coin tossing task. At the end of the experiment, participants returned the sealed envelopes by putting them into a box placed in the corner of the room. Table A3 in the appendix presents descriptive statistics and the randomization check for the placebo experiment. There are no significant differences for any of the elicited background characteristics between the two treatment groups, suggesting that the randomization was successful.

The results show that the general population reported, on average, 56 percent of heads in the coin tossing task (95% confidence interval: [53%, 59%]). This corresponds to 12 percent of misreported coin flips. Thus, participants from the general population cheated too, but to a lesser extent than the criminals ($p < 0.004$, rank-sum test). A clear difference between the sample of criminals and the non-criminal population is the occurrence of the payoff maximizing outcome. While the criminals in the control treatment reported ten times heads in 13 percent of cases, the same outcome is observed only in 4 percent of cases in the control group of the non-criminal population ($p = 0.031$, rank-sum test). However, these differences should

be interpreted cautiously, because they could also be attributed to unobserved factors which differ between the two social groups. The relevant question is whether the treatment influenced cheating within the general population.

Figure 4: Crime-related reminders and cheating in regular citizens



Panel A of this figure shows the distribution of heads reported by the non-criminal population in the control treatment and the binomial distribution implied by honest reporting. Panel B depicts the distribution of heads in the criminal identity treatment and the binomial distribution. Panel C compares the average percentage of heads reported in the control and criminal identity treatment.

Panel A of Figure 4 contrasts the binomial distribution of heads with the empirical distribution observed in the control group of the general population. The distribution is shifted to the right, suggesting that individuals over-reported the number of heads. In contrast to the experiment with the prisoners, the distribution from the criminal profile treatment is very similar to that in the control treatment, as shown in panel B. Panel C highlights that the average percentage of heads is even slightly higher in the control (57 percent of heads or 14 percent misreporting) than in the

criminal profile treatment (55 percent heads or 10 percent misreporting), but not significantly so ($p = 0.240$, rank-sum test). Thus, if anything, the general population tended to cheat less when primed with the criminal profile rather than the control profile. Overall the results from the placebo experiment suggest that the priming effect is specific to criminals and is thus consistent with a criminal identity effect. This is summarized in the final result:

Result 4: *Regular citizens do not cheat more when crime is rendered more salient.*

4 Conclusion

We study the behavioral impact of criminal identity on cheating in 182 inmates from a maximum security prison. We experimentally manipulated the saliency of their criminal identity and subsequently measured their dishonesty in an incentivized task. Our results show that the prisoners cheated substantially more when we rendered their criminal identity more salient. This effect is specific to individuals who possess a criminal identity, as we did not find any effect of crime-related reminders in a placebo experiment with regular citizens. Furthermore, we show that our experimental measure of dishonesty correlates with inmates' disciplinary offenses. Altogether, these findings highlight that criminal identity might play a crucial role in rule violating behavior and therefore support recent theoretical endeavours incorporating identity into economic models of decision making (see Akerlof and Kranton 2000, 2008, Fang and Loury 2005, and Bénabou and Tirole 2011).

From a policy perspective, our results suggest that re-offences could be attenuated by gearing the organization of everyday life in prisons as well as therapy programs to prevent inmates from adopting a criminal identity. Greater interaction with community volunteers, for example, would expose inmates to alternative social identities conducive to law abiding behavior (Wormith 1984, Castleton and Cid 2012). Moreover, our results suggest that it is important for policy makers and legal practitioners to consider the potential side-effects of specific legal institutions and practices

on criminal identity. For example, convictions often trigger additional collateral sanctions for ex-convicts such as deprivations of civil rights, loss of professional licenses, or restricted access to public benefits (Travis 2002). Many of those collateral sanctions could provoke social exclusion, which may strengthen the offenders' criminal identity. Identifying the extent to which the saliency of criminal identity counteracts the deterrent effects of such legal practices is therefore an important avenue for future research.

References

Abeler, Johannes, Anke Becker, and Armin Falk (2012): "Truth-Telling: A Representative Assessment", IZA Discussion Paper No. 6919.

Aizer, Anna and Joseph J. Jr. Doyle (2013): "Juvenile Incarceration, Human Capital and Future Crime: Evidence from Randomly-Assigned Judges", NBER Working Paper No. 19102.

Akerlof, A., George and E. Kranton, Rachel (2008): "Identity, Supervision, and Work Groups", *American Economic Review*, Vol. 98, No. 2, pp. 212–217.

Akerlof, George A. and Rachel E. Kranton (2000): "Economics and Identity", *Quarterly Journal of Economics*, Vol. 115, No. 3, pp. 715–753.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (2013): "On the Origins of Gender Roles: Women and the Plough", *Quarterly Journal of Economics*, Vol. 128, No. 2, pp. 469–530.

Bargh, John A. and Tanya L. Chartrand (2000): "The Mind in the Middle: A Practical Guide to Priming and Automaticity Research", in: H. T. Reis and C. M. Judd (Eds.), *Handbook of Research Methods in Social Psychology*, pp. 253–285, New York: Cambridge University Press.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen (2009): "Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections", *Quarterly Journal of Economics*, Vol. 124, No. 1, pp. 105–147.

Becker, Gary S. (1968): "Crime and Punishment: An Economic Approach", *Journal of Political Economy*, Vol. 76, No. 2, pp. 169–217.

Bénabou, Roland and Jean Tirole (2011): "Identity, Morals, and Taboos: Beliefs as Assets", *Quarterly Journal of Economics*, Vol. 126, No. 2, pp. 805–855.

Benjamin, Daniel J., James J. Choi, and Geoffrey Fisher (2013): "Religious Identity and Economic Behavior", Working paper.

Benjamin, Daniel J., James J. Choi, and Joshua A. Strickland (2010): "Social Identity and Preferences", *American Economic Review*, Vol. 100, No. 4, pp. 1913–1928.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan (2013): "Gender Identitiy and Relative Income within Housholds", Working paper.

Birkeland, Sigbjørn, Alexander W. Cappelen, Erik Ø. Sørensen, and Bertil Tungodden (forthcoming): "An Experimental Study of Prosocial Motivation Among Criminals", *Experimental Economics*.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (forthcoming): "Salience and Consumer Choice", *Journal of Political Economy*.

Bradley, Margaret M. and Peter J. Lang (1994): "Measuring emotion: the self-assessment manikin and the semantic differential", *Journal of Behavior Therapy and Experimental Psychiatry*, Vol. 25, No. 1, pp. 49–59.

Bradley, Margaret M. and Peter J. Lang (2000): "Measuring Emotion: Behavior, Feeling, and Physiology", in: Richard Lane, Lynn Nadel, and Geoffrey Ahern (Eds.), *Cognitive Neuroscience of Emotion*, pp. 242–276, Oxford University Press.

Buonanno, Paolo and Steven Raphael (forthcoming): "Incarceration and Incapacitation: Evidence from the 2006 Italian Collective Pardon", *American Economic Review*.

Bushong, Benjamin, Lindsay M King, Colin F Camerer, and Antonio Rangel (2010): "Pavlovian Processes in Consumer Choice: The Physical Presence of a Good Increases Willingness-to-Pay", *American Economic Review*, Vol. 100, No. 4, pp. 1556–1571.

Callen, Michael, Mohammad Isaqzadeh, James D. Long, and Charles Sprenger (forthcoming): "Violent Trauma and Risk Preference: Experimental Evidence from Afghanistan", *Quarterly Journal of Economics*.

Castleton, Alexander and Alejandro Cid (2012): "Mechanisms Behind Substance Abuse and Rugby: Lessons from a Field Experiment with Incarcerated Offenders", *Desarrollo y Sociedad*, , No. 69, pp. 73–97.

Charness, Gary and Martin Dufwenberg (2006): "Promises and Partnership", *Econometrica*, Vol. 74, No. 6, pp. 1579–1601.

Charness, Gary, Luca Rigotti, and Aldo Rustichini (2007): "Individual Behavior and Group Membership", *American Economic Review*, Vol. 97, No. 4, pp. 1340–1352.

Chen, M., Keith and Jesse M. Shapiro (2007): "Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-based Approach", *American Law and Economics Review*, Vol. 9, No. 1, pp. 1–29.

Chen, Yan and Sherry X. Li (2009): "Group Identity and Social Preferences", *American Economic Review*, Vol. 99, No. 1, pp. 431–457.

Chen, Yan, Sherry X. Li, Tracy X. Liu, and Margaret Shih (2010): "Which Hat to Wear? Impact of Natural Identities on Coordination and Cooperation", Working paper.

Chetty, Raj, Adam Looney, and Kory Kroft (2009): Salience and Taxation: Theory and Evidence, *American Economic Review*, Vol. 99, No. 4, pp. 1145–1177.

Chmura, Thorsten, Christoph Engel, and Markus Englerth (2013): "Selfishness As a Potential Cause of Crime. A Prison Experiment", Working paper.

Clemmer, Donald (1940): *The Prison Community*, Boston: Christopher Publishing House.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011): "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences", *Journal of the European Economic Association*, Vol. 9, No. 3, pp. 522–550.

Drago, Francesco, Roberto Galbiati, and Pietro Vertova (2009): "The Deterrent Effects of Prison: Evidence from a Natural Experiment", *Journal of Political Economy*, Vol. 117, No. 2, pp. 257–280.

Drago, Francesco, Roberto Galbiati, and Pietro Vertova (2011): "Prison Conditions and Recidivism", *American Law and Economics Review*, Vol. 13, No. 1, pp. 103–130.

Ellingsen, Tore and Magnus Johannesson (2004): "Promises, Threats and Fairness", *Economic Journal*, Vol. 114, No. 495, pp. 397–420.

Falk, Armin, Gari Walkowitz, and Wolfgang Wirth (2009): "Benachteiligung wegen mangelnden Vertrauens?: Eine experimentelle Studie zur Arbeitsmarktintegration von Strafgefangenen", *Monatsschrift für Kriminologie und Strafrecht*, Vol. 92, No. 6, pp. 526–546.

Fang, Hanming and Glenn C. Loury (2005): "Dysfunctional Identities Can Be Rational", *American Economic Review*, Vol. 95, No. 2, pp. 104–111.

Fischbacher, Urs and Franziska Föllmi-Heusi (2013): "Lies in Disguise - An Experimental Study on Cheating", *Journal of the European Economic Association*, Vol. 11, No. 3, pp. 525–547.

Frederick, Shane (2005): "Cognitive Reflection and Decision Making", *Journal of Economic Perspectives*, Vol. 19, No. 4, pp. 25–42.

Gabaix, Xavier (2013): "A Sparsity-Based Model of Bounded Rationality, Applied to Basic Consumer and Equilibrium Theory", Working paper.

Gendreau, P., T. Little, and C. Goggin (1996): "Meta-Analysis of the Predictors of Adult Offender Recidivism: What Works!", *Criminology*, Vol. 34, No. 4, pp. 575–607.

Gneezy, Uri (2005): "Deception: The role of consequences", *American Economic Review*, Vol. 95, No. 1, pp. 384–394.

Goette, Lorenz, David Huffman, and Stephan Meier (2012): "The Impact of Social Ties on Group Interactions: Evidence from Minimal Groups and Randomly Assigned Real Groups", *American Economic Journal: Microeconomics*, Vol. 4, No. 1, pp. 101–115.

Gumperz, John J. (1982): *Language and Social Identity*, Vol. 2, Cambridge: Cambridge University Press.

Helland, Eric and Alexander Tabarrok (2007): "Does Three Strikes Deter? A Nonparametric Estimation, *Journal of Human Resources*, Vol. 42, No. 2, pp. 309–330.

Hjalmarsson, Randi (2009): "Juvenile Jails: A Path to the Straight and Narrow or to Hardened Criminality?", *Journal of Law and Economics*, Vol. 52, No. 4, pp. 779–809.

Hoff, Karla and Priyanka Pandey (2006): "Discrimination, Social Identity, and Durable Inequalities", *American Economic Review*, Vol. 96, No. 2, pp. 206–211.

Hoff, Karla and Priyanka Pandey (forthcoming): "Making Up People - The Effect of Identity on Performance in a Modernizing Society", *Journal of Development Economics*.

Houser, Daniel, Stefan Vetter, and Joachim Winter (2012): "Fairness and Cheating", *European Economic Review*, Vol. 56, No. 8, pp. 1645–1655.

Kaminski, Marek (2004): *Games Prisoners Play: The Tragicomic Worlds of Polish Prison*, Princeton: University Press Princeton.

Kessler, Daniel and Steven D. Levitt (1999): "Using Sentence Enhancements to Distinguish Between Deterrence and Incapacitation", *Journal of Law and Economics*, Vol. 42, No. S1, pp. 343–364.

Khadjavi, Menusch and Andreas Lange (2013): "Prisoners and their Dilemma", *Journal of Economic Behavior & Organization*, Vol. 92, pp. 163–175, ISSN 0167-2681.

Killias, Martin, Marcelo Aebi, and Denis Ribeaud (2000): "Does Community Service Rehabilitate better than Short-term Imprisonment?: Results of a Controlled Experiment", *Howard Journal of Criminal Justice*, Vol. 39, No. 1, pp. 40–57.

Kranton, Rachel E., Matthew Pease, Seth Sanders, and Scott Huettel (2012): "Identity, Group Conflict, and Social Preferences", Working paper.

Kuziemko, Ilyana (2013): "How Should Inmates be Released from Prison? An Assessment of Parole versus Fixed-sentence Regimes", *Quarterly Journal of Economics*, Vol. 128, No. 1, pp. 371–424.

Lacetera, Nicola, G. Pope, Devin, and R. Sydnor, Justin (2012): "Heuristic Thinking and Limited Attention in the Car Market", *American Economic Review*, Vol. 102, No. 5, pp. 2206–2236.

LeBoeuf, Robyn A., Eldar Shafir, and Julia Belyavsky Bayuk (2010): "The Conflicting Choices of Alternating Selves", *Organizational Behavior and Human Decision Processes*, Vol. 111, No. 1, pp. 48–61.

Lee, David S. and Justin McCrary (2009): "The Deterrence Effect of Prison: Dynamic Theory and Evidence", Industrial Relations Section Working Paper.

Lerman, Amy E. (2009): "The People Prisons Make: Effects of Incarceration on Criminal Psychology", in: Raphael Steven and Michael Stoll (Eds.), *Do Prisons Make Us Safer?: The Benefits and Costs of the Prison Boom*, pp. 151–176, New York: Russell Sage Foundation.

Levitt, Steven D. (1998): "Juvenile Crime and Punishment", *Journal of Political Economy*, Vol. 106, No. 6, pp. 1156–1185.

Levitt, Steven D. (2002): "Deterrence", in: James Q. Wilson and Joan Petersilia (Eds.), *Crime: Public Policies for Crime Control*, pp. 435–450, Oakland, CA: Institute for Contemporary Studies Press.

Lochner, Lance (2007): "Individual Perceptions of the Criminal Justice System", *American Economic Review*, Vol. 97, No. 1, pp. 444–460.

Manski, Charles F. and Daniel S. Nagin (1998): "Bounding Disagreements about Treatment Effects: A Case Study of Sentencing and Recidivism", *Sociological Methodology*, Vol. 28, No. 1, pp. 99–137.

Mazar, Nina, On Amir, and Dan Ariely (2008): "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance", *Journal of Marketing Research*, Vol. 45, No. 6, pp. 633–644.

Nagin, Daniel S., Francis T. Cullen, and Cheryl Lero Jonson (2009): "Imprisonment and Reoffending", *Crime and Justice*, Vol. 38, No. 1, pp. 115–200.

Pager, Devah (2003): "The Mark of a Criminal Record", *American Journal of Sociology*, Vol. 108, No. 5, pp. 937–975.

Pew Center on the States (2011): *State of Recidivism: The Revolving Door of America's Prisons*, Washington, DC: Pew Charitable Trust.

Pruckner, Gerald J. and Rupert Sausgruber (2013): "Honesty on the Streets: A Field Study on Newspaper Purchasing", *Journal of the European Economic Association*, Vol. 11, No. 3, pp. 661–679.

Shalvi, Shaul, Jason Dana, Michel J.J. Handgraaf, and Carsten DeDreu (2011): "Justified Ethicality: Observing Desired Counterfactuals Modifies Ethical Perceptions and Behavior", *Organizational Behavior and Human Decision Processes*, Vol. 115, No. 2, pp. 181–190.

Sherman, Lawrence W. (1993): "Defiance, Deterrence, and Irrelevance: A Theory of the Criminal Sanction", *Journal of Research in Crime and Delinquency*, Vol. 30, No. 4, pp. 445–473.

Shih, Margaret, Todd L. Pittinsky, and Nalini Ambady (1999): "Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance", *Psychological Science*, Vol. 10, No. 1, pp. 80–83.

Smith, Douglas A. and Patrick R. Gartin (1989): "Specifying Specific Deterrence: The Influence of Arrest on Future Criminal Activity", *American Sociological Review*, Vol. 54, No. 1, pp. 94–106.

Sykes, Gresham M. (1958): *The Society of Captives: A Study of a Maximum Security Prison*, Princeton: University Press Princeton.

Tannenbaum, Frank (1957): *Crime and the Community*, New York: Columbia University Press.

Travis, Jeremy (2002): "Invisible Punishment: An Instrument of Social Exclusion", in: Meda Chesney-Lind and Marc Mauer (Eds.), *Invisible Punishment: The Collateral Consequences of Mass Imprisonment*, New York: The New Press.

Turner, John C. (1985): "Social Categorization and the Self-concept: A Social Cognitive Theory of Group Behavior", in: Edward J. Lawler (Ed.), *Advances in Group Processes*, Vol. 2, pp. 77–122, Greenwich, CT: JAI Press.

Walters, Glenn D. (2003): "Changes in Criminal Thinking and Identity in Novice and Experienced Inmates Prisonization Revisited", *Criminal Justice and Behavior*, Vol. 30, No. 4, pp. 399–421.

Wamsley, Roy (2011): *World Prison Polulation List (Ninth Edition)*', Kings's Collegge London, International Center for Prison Studies.

Western, Bruce, Jeffrey R. Kling, and David F. Weiman (2001): "The Labor Market Consequences of Incarceration", *Crime & Delinquency*, Vol. 47, No. 3, pp. 410–427.

Wormith, Stephen J. (1984): "Attitude and Behavior Change of Correctional Clientele", *Criminology*, Vol. 22, No. 4, pp. 595–618.

Appendix

A Additional Tables

Table A1: Descriptive statistics for the main experiment

	Total sample N = 182		**Criminal identity** N = 90		**Control** N = 92		
Variable	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**p-value**
Type of conviction:							
Violent crimes	0.297	(0.458)	0.256	(0.439)	0.337	(0.475)	0.229
Drug-related crimes	0.264	(0.442)	0.278	(0.450)	0.250	(0.435)	0.671
Property crimes	0.242	(0.429)	0.267	(0.445)	0.217	(0.415)	0.438
Sex crimes	0.148	(0.356)	0.189	(0.394)	0.109	(0.313)	0.128
Other	0.049	(0.217)	0.011	(0.105)	0.087	(0.283)	0.018
Repeat offender	0.621	(0.487)	0.567	(0.498)	0.674	(0.471)	0.136
Sentences:							
Sentence length (if known)	4.574	(4.364)	5.082	(4.580)	4.065	(4.114)	0.167
Safe custody	0.176	(0.382)	0.178	(0.384)	0.174	(0.381)	0.945
Early imprisonment	0.176	(0.382)	0.167	(0.375)	0.185	(0.390)	0.748
Years in prison	2.659	(3.922)	3.055	(4.433)	2.272	(3.326)	0.100
Prison section:							
Double cell	0.352	(0.479)	0.367	(0.485)	0.337	(0.475)	0.675
Single cell (normal)	0.368	(0.484)	0.356	(0.481)	0.380	(0.488)	0.728
Single cell (special)	0.280	(0.450)	0.278	(0.450)	0.283	(0.453)	0.942
No. of disciplinary offenses	2.187	(4.371)	2.533	(4.432)	1.848	(4.307)	0.131
Nationality:							
Swiss	0.322	(0.468)	0.369	(0.485)	0.276	(0.450)	0.192
South-eastern European	0.298	(0.459)	0.262	(0.442)	0.333	(0.474)	0.307
African	0.146	(0.354)	0.143	(0.352)	0.149	(0.359)	0.903
Central European	0.123	(0.329)	0.107	(0.311)	0.138	(0.347)	0.540
Other	0.111	(0.315)	0.119	(0.326)	0.103	(0.306)	0.746
Age	38.341	(11.306)	39.246	(12.059)	37.497	(10.553)	0.479
Highest completed education:							
Compulsory school	0.376	(0.486)	0.437	(0.499)	0.314	(0.467)	0.095
Vocational school	0.347	(0.477)	0.345	(0.478)	0.349	(0.479)	0.956
High school	0.087	(0.282)	0.057	(0.234)	0.116	(0.322)	0.169
Teaching diploma	0.023	(0.151)	0.011	(0.107)	0.035	(0.185)	0.306
Adv. vocational school	0.104	(0.306)	0.115	(0.321)	0.093	(0.292)	0.637
Univ. of applied sciences	0.035	(0.184)	0.023	(0.151)	0.047	(0.212)	0.398
University	0.029	(0.168)	0.011	(0.107)	0.047	(0.212)	0.169
Risk attitudes	5.392	(3.036)	5.163	(2.854)	5.626	(3.212)	0.294
Cognitive skills	0.676	(0.974)	0.778	(1.036)	0.576	(0.905)	0.152
Survey language:							
German	0.709	(0.456)	0.722	(0.450)	0.696	(0.463)	0.693
English	0.126	(0.333)	0.111	(0.316)	0.141	(0.350)	0.540
Italian	0.099	(0.299)	0.089	(0.286)	0.109	(0.313)	0.655
French	0.066	(0.249)	0.078	(0.269)	0.054	(0.228)	0.524

Descriptive statistics for the main experiment. All variables are binary, except for sentence length (in years), years in prison, number of disciplinary offenses, age (in years), risk attitudes (ranging from 0 ''not at all willing to take risks" to 10 ''fully willing to take risks"), and cognitive skills (score in the cognitive reflection test ranging from 0 to 3). ''Single cell (normal)" and ''Single cell (special)" means single cell in normal, respectively special correction facility. Sentence length is known for 118 subjects. Due to item non-response, 11 observations are missing for nationality, 8 for age, 9 for education, and 6 for risk attitudes. The last column presents p-values for the null hypothesis of perfect randomization (X^2 tests in case of binary variables and rank-sum tests in case of interval variables).

Table A2: Descriptive statistics for the manipulation check

	Total sample N = 119		**Criminal identity** N = 52		**Control** N = 67		
Variable	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**p-value**
Type of conviction:							
Violent crimes	0.319	(0.468)	0.346	(0.480)	0.299	(0.461)	0.580
Drug-related crimes	0.269	(0.445)	0.327	(0.474)	0.224	(0.420)	0.209
Property crimes	0.218	(0.415)	0.192	(0.398)	0.239	(0.430)	0.543
Sex crimes	0.151	(0.360)	0.096	(0.298)	0.194	(0.398)	0.139
Other	0.042	(0.201)	0.038	(0.194)	0.045	(0.208)	0.865
Repeat offender	0.538	(0.501)	0.519	(0.505)	0.552	(0.501)	0.720
Sentences:							
Sentence length (if known)	5.649	(4.260)	4.674	(3.295)	6.327	(4.746)	0.131
Safe custody	0.218	(0.415)	0.250	(0.437)	0.194	(0.398)	0.464
Early imprisonment	0.269	(0.445)	0.269	(0.448)	0.269	(0.447)	0.994
Years in prison	2.522	(4.005)	2.246	(3.787)	2.736	(4.181)	0.183
Prison section:							
Double cell	0.303	(0.461)	0.308	(0.466)	0.299	(0.461)	0.914
Single cell (normal)	0.319	(0.468)	0.250	(0.437)	0.373	(0.487)	0.153
Single cell (special)	0.378	(0.487)	0.442	(0.502)	0.328	(0.473)	0.204
No. of disciplinary offenses	2.571	(5.148)	2.442	(5.707)	2.672	(4.711)	0.226
Survey language:							
German	0.840	(0.368)	0.808	(0.398)	0.866	(0.344)	0.392
English	0.084	(0.279)	0.096	(0.298)	0.075	(0.265)	0.675
Italian	0.050	(0.220)	0.077	(0.269)	0.030	(0.171)	0.244
French	0.025	(0.157)	0.019	(0.139)	0.030	(0.171)	0.714

Descriptive statistics for the manipulation check. All variables are binary, except for sentence length (in years), years in prison, and number of disciplinary offenses. ''Single cell (normal)" and ''Single cell (special)" means single cell in normal, respectively special correction facility. Sentence length is known for 61 subjects. The last column presents p-values for the null hypothesis of perfect randomization (X^2 tests in case of binary variables and rank-sum tests in case of interval variables).

Table A3: Descriptive statistics for the placebo experiment

	Total sample N = 193		**Crime prime** N = 98		**Control** N = 95		
Variable	**mean**	**sd**	**mean**	**sd**	**mean**	**sd**	**p-value**
Nationality:							
Swiss	0.587	(0.494)	0.573	(0.497)	0.602	(0.492)	0.683
South-eastern European	0.106	(0.308)	0.083	(0.278)	0.129	(0.337)	0.307
African	0.011	(0.103)	0.010	(0.102)	0.011	(0.104)	0.982
Central European	0.243	(0.430)	0.271	(0.447)	0.215	(0.413)	0.372
Other	0.053	(0.224)	0.063	(0.243)	0.043	(0.204)	0.550
Age	41.605	(17.155)	42.930	(18.236)	40.251	(15.963)	0.424
Highest completed education:							
Compulsory school	0.095	(0.294)	0.126	(0.334)	0.064	(0.246)	0.143
Vocational school	0.365	(0.483)	0.295	(0.458)	0.436	(0.499)	0.043
High school	0.106	(0.308)	0.137	(0.346)	0.074	(0.264)	0.163
Teaching diploma	0.037	(0.189)	0.021	(0.144)	0.053	(0.226)	0.242
Adv. vocational school	0.175	(0.381)	0.158	(0.367)	0.191	(0.396)	0.543
Univ. of applied sciences	0.095	(0.294)	0.116	(0.322)	0.074	(0.264)	0.333
University	0.127	(0.334)	0.147	(0.356)	0.106	(0.310)	0.397
Risk attitudes	5.940	(2.423)	5.724	(2.503)	6.165	(2.330)	0.259
Cognitive skills	1.228	(1.186)	1.245	(1.149)	1.211	(1.228)	0.722
Survey language:							
German	0.938	(0.242)	0.929	(0.259)	0.947	(0.224)	0.589
English	0.041	(0.200)	0.041	(0.199)	0.042	(0.202)	0.964
Italian	0.005	(0.072)	0.010	(0.101)	0.000	(0.000)	0.324
French	0.016	(0.124)	0.020	(0.142)	0.011	(0.103)	0.579

Descriptive statistics for the placebo experiment. All variables are binary, except for age (in years), risk attitudes (ranging from 0 ''not at all willing to take risks" to 10 ''fully willing to take risks"), and cognitive skills (score in the cognitive reflection test ranging from 0 to 3). Due to item non-response, 4 observations are missing for nationality and education, 11 for age, 1 for risk attitudes. The last column presents p-values for the null hypothesis of perfect randomization (X^2 tests in case of binary variables and rank-sum tests in case of interval variables).

Table A4: Descriptive statistics for total prison population based on the annual report.

Variable	**Total (N = 422)**	**in %**
Nationality:		
Swiss	111	26.30%
South-eastern European	120	28.44%
African	72	17.06%
Central European	45	10.66%
Other	74	17.54%
Age:		
29 years or younger	122	29%
30-39 years	126	30%
40-49 years	97	23%
50-59 years	62	15%
60 years or older	15	4%
Sentence length (if known):		
less than 6 months	19	5%
6-12 months	3	1%
1-2 years	8	2%
2-3 years	37	9%
3-5 years	69	16%
5-10 years	87	21%
10-20 years	43	10%
more than 20 years	4	1%
Safe custody	82	19%
Early imprisonment	70	17%
Type of conviction:		
Violent crimes	121	29%
Drug-related crimes	118	28%
Property crimes	79	19%
Sex crimes	56	13%
Other	48	11%