Review of Economic Studies (2015) **82**, 1289–1308 doi:10.1093/restud/rdv025
© The Author 2015. Published by Oxford University Press on behalf of The Review of Economic Studies Limited.
Advance access publication 23 June 2015

Bad Boys: How Criminal Identity Salience Affects Rule Violation

ALAIN COHN
Booth School of Business, University of Chicago
MICHEL ANDRÉ MARÉCHAL
Department of Economics, University of Zurich
and
THOMAS NOLL
Swiss Prison Staff Training Center

First version received February 2014*; final version accepted May* 2015 (*Eds.*)

We conducted an experiment with 182 inmates from a maximum security prison to analyze the impact of criminal identity salience on cheating. The results show that inmates cheat more when we exogenously render their criminal identity more salient. This effect is specific to individuals who have 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' offences against in-prison regulation. Together, these findings suggest that criminal identity salience plays a crucial role in rule violating behaviour.

Key words: Dishonesty, Identity, Crime, Prison, Experiment

JEL Codes: K00, C93, K14, K42, Z10

1. INTRODUCTION

Crime imposes substantial costs on society. For example, it has been estimated that the annual burden of crime for the United States ranges between one and two trillion US dollars (Anderson, 1999; Ludwig, 2006). Understanding why people commit crime and violate the law is, therefore, of key interest for policy makers and researchers. According to the seminal work of Becker (1968), a person's decision to commit crime is based on weighing the expected material benefits against the punishment costs.[1] A long-standing sociological and psychological literature argues, however, that in addition to rational cost–benefit calculations, the decision to commit or refrain from crime also crucially depends on people's moral and self-image considerations (see Cohen,

1. See Draca and Machin (2015), Chalfin and McCrary (2014), or Nagin (2013) or for excellent reviews of the empirical literature.

1955; Becker, 1963; Reckless, 1967; Kaplan, 1975; Anderson, 2000).[2] These views line up with recent theoretical approaches in law and economics incorporating identity and norms into decision making (see Akerlof and Kranton, 2000; McAdams and Rasmusen, 2007; Posner, 2009; Bénabou and Tirole, 2011a,b; Acemoglu and Jackson, 2014). For example, people in Akerlof and Kranton's model have multiple social identities (*e.g.* based on gender, ethnicity, or occupation) that are tied to norms prescribing how one should behave. Identities may influence behaviour because individuals experience disutility if their behaviour deviates from what their identities prescribe.

Applying identity theory to the context of crime, we propose that deviant people have a moral and a criminal identity; these identities differ in the extent to which they impose rule compliance. Rule violating behaviour inflicts psychological costs on the moral identity (*e.g.* Gneezy, 2005; Kartik, 2009; Abeler *et al.*, 2014). Whenever these costs are sufficiently large relative to the net benefit of rule violation, the moral identity prescribes compliance. In contrast, the criminal identity does not suffer from moral costs (or suffers to a much lesser extent) and thus prescribes, in accordance with Becker's model, breaking rules whenever it pays off. As a consequence, stronger criminal identification should lead to more rule violating behaviour. While several correlational studies from criminology indeed report a positive relationship between survey measures of criminal identity and deviant behaviour (see Gendreau *et al.*, 1996 for a meta-analysis), these studies are not able to disentangle whether criminal identity has a causal impact on rule violating behaviour or whether rule violating behaviour leads to stronger criminal identification.

Identifying the causal influence of criminal identity on rule violating behaviour is challenging. A simple comparison of criminals and non-criminals is problematic because criminals and law-abiding citizens may differ in many dimensions that are difficult to control for, such as financial background, life prospects, or opportunity costs.[3] Instead, we experimentally manipulate the saliency of criminal identity, which circumvents the aforementioned problems involved in the comparison of people with different levels of criminal identification. Identity theory posits that the extent to which individuals' criminal and moral identities guide their behaviour depends on the relative weight (*i.e.* salience) they attach to each identity at a given moment (see also Benjamin *et al.*, 2010, 2015). Consequently, a temporary boost in criminal identity salience reveals the identity's marginal behavioural effect.

To test whether criminal identity salience triggers rule violating behaviour, we conducted an experiment with 182 inmates from the maximum security prison Pöschwies, Switzerland's largest penitentiary for male adults. We opted for this unusual subject pool because prison inmates have plausibly a distinct criminal self-concept due to their extensive histories of criminal misconduct (*e.g.* see Walters, 2003 or Lerman, 2009). In our experiment, we randomly increased 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 of?"). 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?"). The participants subsequently performed a simple coin tossing task, which allows us to measure their dishonesty as a form of rule violating behaviour. The rules of the task required subjects to flip 10 coins and report the outcomes 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

2. These theories initiated an active literature in criminology developing psychological scales of criminal self-concept (*e.g.* Simourd and Olver, 2002).

3. See Birkeland *et al.* (2014), Khadjavi and Lange (2013), and Chmura *et al.* (2013) for studies comparing distributive preferences and cooperation in criminals and non-criminals.

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.[4]

The results show that a significant share of the inmates cheated. On average, they reported heads for 60% of the coin flips in the control condition. This is significantly above chance and approximates 20% of coin flips being misreported.[5] However, inmates became even more dishonest when their criminal identity was rendered more salient. In the criminal identity treatment, they reported 66% of heads, which corresponds to 32% of coin flips being misreported. Thus, the higher saliency of criminal identity increased the frequency of misreporting by 60%. Using administrative data, we further show that behaviour in the coin tossing task correlates with inmates' offences against in-prison regulation (*e.g.* aggression against others, use of illegal drugs, or weapon possession), suggesting that the coin tossing task provides an externally valid measure of rule violating behaviour. 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 negative emotions and arousal because they reminded prisoners about their deeds or the social injustice of incarceration (Sherman, 1993). Such an emotional reaction could potentially undermine honesty. However, we demonstrate that the criminal identity questions did not influence emotions, and that the correlations between emotions and cheating are insignificant. 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 priming effect.

Finally, the mere exposure to crime-related content, rather than the salience of criminal identity, could have triggered dishonest behaviour in the criminal identity treatment. We, therefore, conducted an additional placebo experiment with "regular" citizens to test whether the priming effect is specific to people who have a criminal identity. We recruited 193 male citizens from the general population and administered the same survey. 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. Subjects subsequently 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% in the control condition to 10% in the criminal condition. Together, the three studies suggest that criminal identity salience promotes rule violating behaviour.

Our results contribute to a growing literature studying the role of identity in economic decision making (*e.g.* Akerlof and Kranton, 2000; Fang and Loury, 2005; or Bénabou and Tirole, 2011a). The empirical literature mostly analysed whether people discriminate between in- and outgroup

4. See Fischbacher and Föllmi-Heusi (2013) or Houser *et al.* (2012) for similar approaches to elicit dishonest behaviour.

5. In line with previous research on honesty, our subjects did not take full advantage of cheating opportunities (Ariely, 2012).

members (*e.g.* Hoff and Pandey, 2006; Charness *et al.*, 2007; Chen and Li, 2009; Goette *et al.*, 2012; or Kranton *et al.*, 2013). One of the few exceptions is the study by Benjamin *et al.* (2010) who used a similar approach to ours and analysed the influence of ethnic and gender identity salience on risk and time preferences. A more recent study by Bertrand *et al.* (forthcoming) illustrates that gender identity norms influence a wide range of economic and social outcomes. They find that the norm "wives should not earn more than their husbands" helps explain female labour force participation, divorce, and the division of home production within U.S. households.[6] We add to this literature by providing causal evidence on the impact of identity salience on dishonest behaviour. In this sense, our results are also relevant for the rapidly expanding literature on the determinants of dishonesty (*e.g.* Gneezy, 2005; Mazar *et al.*, 2008; Shalvi *et al.*, 2011; Balafoutas *et al.*, 2013; Fischbacher and Föllmi-Heusi, 2013; Pruckner and Sausgruber, 2013; or Cohn *et al.*, 2014).

Our study may also be relevant for the problem of recidivism. Estimates from the U.S., for example, indicate that more than 4 out of 10 released inmates are rearrested within three years (Pew Center on the States, 2011). Our results suggest that one factor that may contribute to recidivism are the so-called collateral consequences of conviction. These are punishments that come in addition to the actual sentence and include deprivation of civil rights, loss of professional licenses, or restricted access to public benefits (Travis, 2002). Besides deterring people, collateral sanctions may increase the probability of recidivism by reminding convicts of their criminal identities.

2. AN EXPERIMENT BEHIND BARS

2.1. *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 (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.[7]

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 available as Supplementary Material), 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.[8] The experiment was conducted within 24 hours to minimize the possibility of cross-talk among inmates.

The first part of the survey contained filler questions on subjective well-being and standard demographics. The second part comprised our key experimental manipulation. We randomly

6. See also Alesina *et al.* (2013) on the historical origins of today's gender identity norms.

7. See Table A.1 in the Appendix available as Supplementary Material for further descriptive statistics of our sample. Table A.4 in the appendix available as Supplementary Material 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.

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

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?").[9] 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 difference between the criminal identity and the control treatment.[10] Immediately after the priming, participants were asked to indicate their current emotional state using non-verbal Self-Assessment Manikins (Bradley and Lang, 1994).[11] 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 10 coins, each worth 0.5 Swiss francs (or 0.55 U.S. 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 sizeable 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 behaviour. 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 coin flips that are misreported (see also Houser *et al.*, 2012). Let h be the percentage of reported heads and m the percentage of coin tosses being misreported. 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 coin tosses that are misreported by

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

The last column of Table A.1 in the appendix available as Supplementary Material 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 offences, re-offending status, age, cognitive skills[12], 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

9. Priming is a method developed in psychology and refers to the activation of mental representations through situational cues (Shih *et al.*, 1999; LeBoeuf *et al.*, 2010; or Bargh and Chartrand, 2000). Priming is now increasingly used in economics (*e.g.* Benjamin *et al.*, 2010, 2015; Chen *et al.*, 2014; Callen *et al.*, 2014; or Hoff and Pandey, 2014).

10. To ensure a ceteris paribus comparison of the two treatments, we also matched the answer formats of the two sets of questions.

11. 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).

12. Cognitive skills were elicited using Frederick's (2005) cognitive reflection test.

($p=0.100$, 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.

2.2. *Framework and Hypotheses*

We derive our hypotheses by adapting the identity framework from Benjamin *et al.* (2010). We assume that inmates have a criminal identity and a moral identity. These identities prescribe different behaviours in the coin tossing task. As a consequence, an internal conflict arises over appropriate behaviour. We begin by formalizing the behaviour prescribed by the moral identity.

For the moral self, we assume that the decision to cheat depends on personal benefits and expected punishment (Becker, 1968), as well as psychological costs of cheating (*e.g.* Ellingsen and Johannesson, 2004; Gneezy, 2005; Charness and Dufwenberg, 2006; Mazar *et al.*, 2008; Kartik, 2009; or Abeler *et al.*, 2014).[13] Let $x_i \in [0,10]$ denote the number of heads reported by inmate i and x_0 be his true number of tossed heads. The moral identity's prescribed action x_m is derived from maximizing the following utility function:

$$\max_{x_i \in [0,10]} EU_i(x_i) = (1-p)x_i - p[\mathbb{1}_{x_i \neq x_0} f - (1-\mathbb{1}_{x_i \neq x_0})x_i] - \lambda(x_i - x_0)^2, \tag{3}$$

where p is the monitoring probability and f the fine imposed when caught cheating, *i.e.* $x_i \neq x_0$. λ denotes the psychological costs of cheating. Given that in our context the probability of monitoring is zero, *i.e.* $p=0$, the decision to cheat reduces to a tradeoff between monetary gains and psychological costs of cheating. Solving the maximization problem yields the moral identity's optimal action $x_m = x_0 + \frac{1}{2\lambda}$, which depends on the true outcome x_0 and the psychological costs of cheating λ. The larger the psychological costs, the more x_m corresponds to the true outcome x_0.

In contrast to the moral identity, we assume that cheating imposes no psychological costs on the criminal identity. Consequently, the action prescribed by the criminal identity x_c is derived from maximizing the following utility function:

$$\max_{x_i \in [0,10]} EU_i(x_i) = (1-p)x_i - p[\mathbb{1}_{x_i \neq x_0} f - (1-\mathbb{1}_{x_i \neq x_0})x_i]. \tag{4}$$

Given a monitoring probability of zero, the utility of the criminal identity is maximized if heads is reported 10 times, *i.e.* $x_c = 10$. Thus, the criminal identity prescribes cheating to the full extent. We further assume that a criminal person maximizes a utility function that is a convex combination of the prescribed actions of his criminal identity and moral identity. Deviating from the prescribed actions causes disutility. A criminal's maximization problem can thus be characterized as follows:

$$\max_{x_i \in [0,10]} U_i = -w(s)(x_i - x_c)^2 - (1-w(s))(x_i - x_m)^2. \tag{5}$$

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. We assume that $w(0)=0$ and $w' > 0$. s is the current strength of the criminal identity and has a permanent component $\bar{s}$, which may differ between individuals. Environmental cues or primes can make the criminal identity temporarily salient by a factor ϵ (*i.e.* $s = \bar{s} * \epsilon$), where ϵ can take any value between 0 and 1. A value of zero means

13. Note that we use a reduced-form representation of lying costs and do not discriminate between whether these costs arise from internal (*e.g.* self-image concerns) or external regulatory processes (*e.g.* social image concerns).

that the criminal identity is not at all salient, whereas a value of one means that the criminal identity is fully salient.[14] Inserting the preferred actions x_c and x_m into equation (5) and solving the maximization problem gives the following optimal action for inmate i:

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

The optimal action of a criminal with identity considerations is the weighted average of the prescribed actions of his criminal identity and moral identity. By inducing an inmate to think about his criminal background in our experiment, we increase the current strength s of his criminal identity by a factor ϵ and, therefore, augment the decision weight attached to his criminal identity. According to equation (6), his optimal action x_i^* shifts towards the action the criminal identity prescribes which is $x_c = 10$. This leads to the following hypothesis:

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

Note that this prediction only applies to individuals who have a criminal identity. If a person does not have a criminal identity (*i.e.* $\bar{s} = 0$), crime-related primes do not lead to more cheating because $s = \bar{s} \cdot \epsilon = 0$. Consequently, our framework predicts that non-criminals will not report more successful coin flips if they are induced to think about crime. We will return to this point when we discuss our placebo experiment with regular citizens.

3. EXPERIMENTAL RESULTS

We outline our results in four steps. First, we examine the impact of criminal identity salience on cheating. Secondly, we validate our experimental measure of cheating by showing that behaviour in the coin tossing task is correlated with inmates' offences against in-prison regulation. Thirdly, we analyse data from an additional experiment, 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 priming effect is specific to criminals.

3.1. *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% of the coin flips, which is significantly above chance (95% confidence interval: 60%, 66%[15]). Assuming that none of the prisoners cheated to his disadvantage—reporting tails when the actual outcome was heads—we estimate that 26% 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 behaviour predicts. For example, while we should theoretically expect around 0.1% of the participants

14. Given that our subjects live behind bars their criminal identity may already be salient to some extent in the control treatment (*i.e.* $\epsilon > 0$). On the other hand, it seems also plausible that our subjects distract themselves and do not think permanently about the fact that they are incarcerated criminals (*i.e.* $\epsilon < 1$ in the control group). The more salient criminal identities are in the control condition, the less likely we will find a treatment effect.

15. The confidence interval is based on individual averages to account for the fact that reporting behaviour could be correlated within individuals.

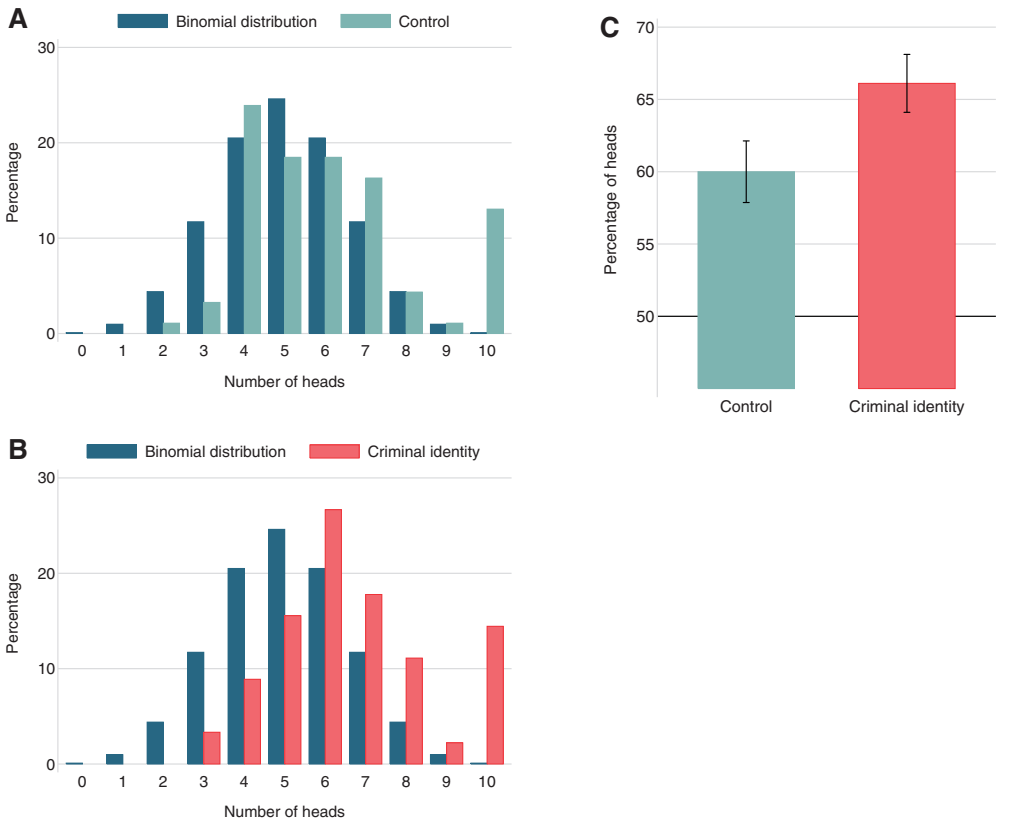

FIGURE 1

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 treatment. Error bars indicate standard error of the mean.

to win the maximum amount, almost 13% of the prisoners 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 the outcomes of 8, 7, and 6 times heads were significantly more frequent than honest reporting would predict (p's < 0.001, 0.007, 0.059, and 0.097, Binomial tests).

As a result, the average percentage of heads increased significantly from 60% in the control group to 66% in the criminal identity group ($p = 0.017$, rank-sum test), as shown in panel C.[16] The corresponding rates of misreported coin flips are 20%, and 32% respectively, suggesting that cheating is 60% 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.

16. All p-values reported in this article are based on two-sided tests except for the Binomial tests, which are based on directed hypotheses.

TABLE 1
Regression analysis: criminal identity and cheating

Dependent variable:	(1)	(2) heads = 1	(3)
Criminal identity	0.061**	0.069**	0.071**
	(0.029)	(0.031)	(0.031)
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

Notes: This table reports average marginal effects from a Probit model. Robust standard errors, corrected for clustering at 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: ** $p<0.05$.

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

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

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 $\mathbf{X}_i$, because these variables were imperfectly balanced across treatments. ϵ_{ik} is the idiosyncratic error term (standard errors are adjusted for clustering at the individual level). $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

The average treatment 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 ($p=0.035$). As shown in column (2), the estimate of the treatment effect is even slightly larger when we include the additional background characteristics as control variables ($p=0.026$).[17] These findings are summarized in the following result:

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

We tested the relevance of 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). However, the results show that the saliency of criminal identity neither had an effect on arousal ($p=0.237$, rank-sum test) nor on negative affect ($p=0.280$, rank-sum test).

Secondly, 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

17. We alternatively estimated a linear probability model using ordinary least squares (OLS) which yielded the same results.

it was impossible to detect whether an 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 before 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 behaviour 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. Column (3) of Table 1 shows that none of these variables reaches statistical significance and the coefficient estimate for the criminal identity treatment remains unchanged ($p = 0.024$). Together, these results suggest that neither emotions nor risk attitudes are able to explain the treatment effect.

Thirdly, it could be that the prisoners were not completely honest when we asked them about their offences to appear less of a bad person. This could have spilled over to the coin tossing task and undermined honest reporting. For example, it may be easier to rationalize dishonest behaviour if one had just lied. If that were the case, it would not be the saliency of criminal identity, but a form of moral disengagement that increased cheating in the criminal identity condition. To test this conjecture, we examined how many prisoners misreported their convictions in the questionnaire (by comparing self-reports to administrative records). Only 7% of the prisoners' answers are inconsistent, and an additional 5% of prisoners failed to answer the question. We re-estimated our main regression model (7) and restricted the sample to inmates who reported their convictions truthfully. If misreporting in the criminal identity questionnaire induced dishonesty in the subsequent coin tossing task we should observe a weaker or no priming effect in the restricted sample. However, the regression results in Table 2 show that the treatment effect remains virtually the same when we exclude inmates who did not provide truthful responses (columns 1 and 2), or additionally also exclude inmates who did not answer the question (columns 3 and 4).[18]

3.2. *Validity of the coin tossing task*

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

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

We regressed the number of offences 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

18. More generally, one could also think of our result as a self-consistency effect, that is, prisoners may want to behave consistently with their recollection of their previous crimes. Such a self-consistency effect is closely related to the concept of identity because individuals often define themselves through their past behaviour (Bem, 1972). Hence, reminding prisoners of their criminal past should make their criminal identity more salient. The idea that individuals consider their past behaviour to infer their identity is also shared by the recent economic literature on identity. For example, Bénabou and Tirole (2011a) propose a model where individuals only imperfectly recall their sense of identity and use past behaviour as a signal of who they are. In a field experiment, Kessler and Milkman (2014) show that reminding charitable donors of their past donations increases current giving, and that the effect is more pronounced for individuals who donate on a regular basis. We thank an anonymous referee for pointing this out.

TABLE 2
Robustness checks: consistency

Dependent variable:	(1)	(2)	(3)	(4)
	heads = 1			
Criminal identity	0.062**	0.072**	0.063**	0.076**
	(0.030)	(0.032)	(0.031)	(0.033)
Additional controls:	no	yes	no	yes
Observations	1690	1600	1600	1510
Subjects	169	160	160	151
Sample: conviction is...	truthful or not reported		truthful	

Notes: This table reports average marginal effects from a Probit model. Robust standard errors, corrected for clustering at the individual level, are displayed in parentheses. The decision to report heads is regressed on the criminal identity treatment dummy. In columns (1) and (2) the sample is restricted to inmates providing truthful (*i.e.* consistent with administrative records) or no information concerning their conviction. In columns (3) and (4) we additionally excluded inmates who did not provide information concerning their conviction from the sample. The additional controls in columns (2) and (4) are the residual category of convictions, the number of years in prison, and compulsory schooling level, because these variables were imperfectly balanced across treatments. Due to item non-response, the number of observations drops when adding covariates. Significance levels: $^{**}p < 0.05$.

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 $\mathbf{X}_i$. ϵ_i is the idiosyncratic error term. Column (1) of Table 3 shows that, on average, inmates who reported ten times heads committed two more offences in prison than those who reported heads in 50% 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 even when we control for criminal background and socio-economic characteristics, as shown in columns (2) and (3).[19] The following result summarizes these findings:

Result 2: *Behaviour in the coin tossing experiment correlates with rule violating behaviour in prison.*

3.3. *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 well-being (see online appendix available as Supplementary Material). The second part contained exactly the same six priming questions as in the previous experiment.[20] Following the six priming questions, participants solved a word stem completion task. For example, they could complete

19. We alternatively treated disciplinary offences as count data and estimated a negative binomial regression model, which yielded similar results.

20. Sixty-nine inmates had already participated in the first experiment; the other 50 inmates participated for the first time. Table A.2 in the appendix available as Supplementary Material 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. Treatments were randomly assigned for subjects who participated for the first time. 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. Seven subjects received the same treatment in the main experiment and the manipulation check by mistake. The results are robust to excluding these subjects.

TABLE 3
Prison rule violations and behaviour in the coin tossing task

Dependent variable:	(1)	(2) # of disciplinary offences	(3)
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

Notes: This table reports OLS coefficient estimates. Robust standard errors are displayed in parentheses. In column (1), the number of disciplinary offences 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: $^{**}p<0.05$, $^{***}p<0.01$.

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 versus account) and "pol..." (*e.g.* police versus politics). A research assistant who was blind to the experimental conditions categorized the answers into crime-related and unrelated words. This allows us to compare the mental accessibility of crime-related constructs across treatments.

As depicted in Figure 2, 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.*

3.4. *Placebo experiment with regular citizens*

We conducted an additional placebo experiment with subjects from the general population to further consolidate our results with an effect of criminal identification. An alternative explanation for the previous findings is that not the salience of criminal identity per se, but rather crime-related thoughts in general, increased cheating. It is conceivable that exposure to crime-related content evoked semantically related thoughts and thereby lowered moral standards. This alternative mechanism would imply that regular citizens react similarly to crime-related priming questions. In contrast, our identity framework predicts that non-criminal subjects will not react to crime-related cues because they do not have a criminal identity that can be made more salient (*i.e.* $\bar{s}=0$ and therefore $s=\bar{s}*\epsilon=0$).[21]

21. See LeBoeuf *et al.* (2010) for evidence that identity priming only affects behaviour if subjects have the primed identity.

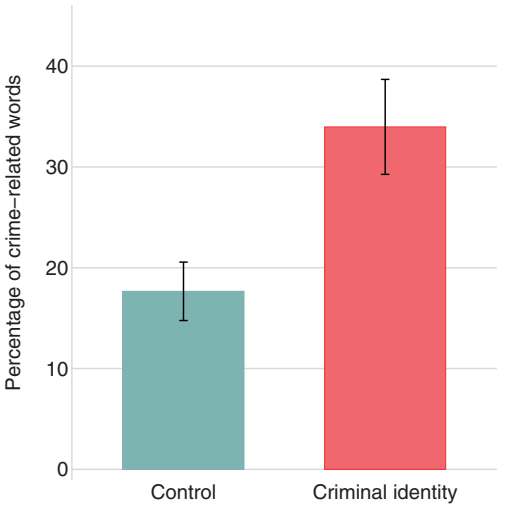

FIGURE 2

Manipulation check. This figure shows the average percentage of crime-related words in the word completion task by treatment. Error bars indicate standard error of the mean.

To test this prediction, we recruited 193 male visitors from the resident registration office of a Swiss municipality.[22] We deliberately chose a municipality characterized by a relatively high proportion of foreigners 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.[23] The 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 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 available as Supplementary Material). Subjects subsequently answered the same six priming questions as the prisoners but 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 A.3 in the appendix available as Supplementary Material

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

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

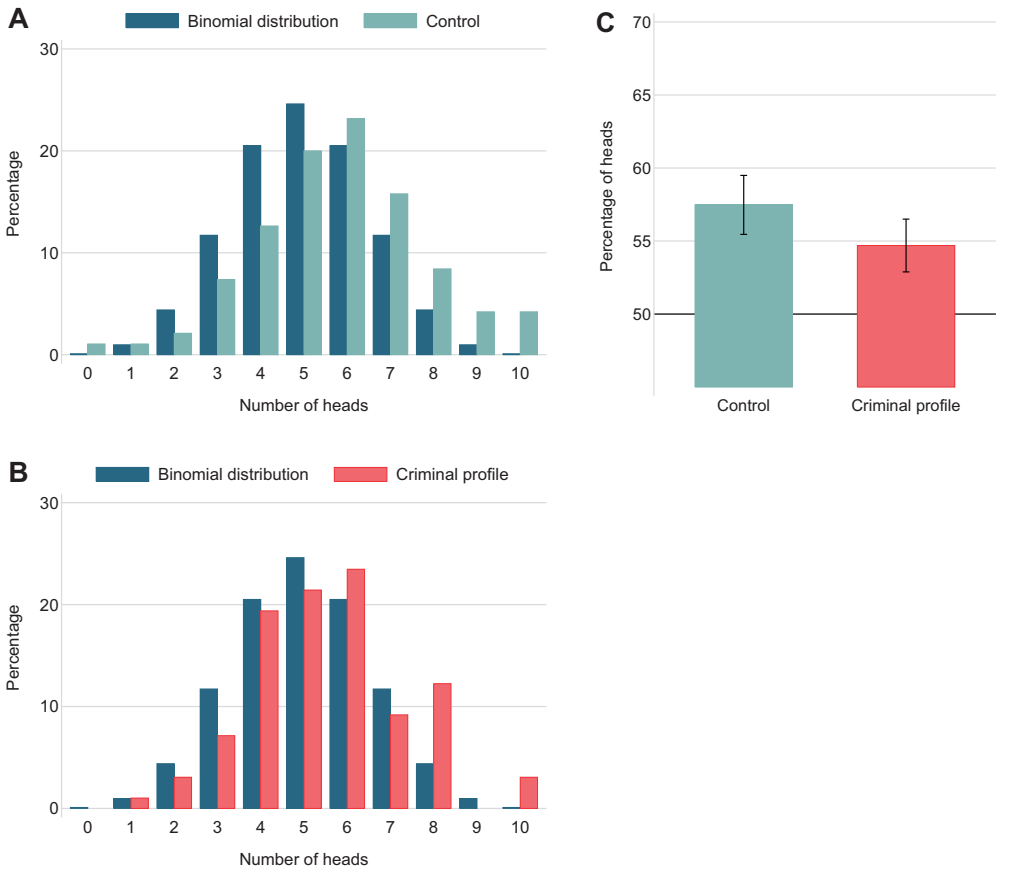

FIGURE 3

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 profile treatment and the binomial distribution. Panel C compares the average percentage of heads reported in the control and criminal profile treatment. Error bars indicate standard error of the mean.

presents descriptive statistics and the randomization check for the placebo experiment. There are no significant differences between treatments for any of the elicited background characteristics, with the exception of vocation schooling ($p=0.043$, χ^2 test), suggesting that the randomization was successful.

The results show that the general population reported, on average, 56% of heads in the coin tossing task (95% confidence interval: 53%, 59%). This corresponds to 12% misreporting. 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 10 times heads in 13% of cases, the same outcome is observed only in 4% of cases in the control group of the non-criminal population ($p=0.031$, χ^2 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.

Panel A of Figure 3 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 lower in the criminal profile (55% heads or 10% misreporting) compared to the control treatment (57% of heads or 14% 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-related concepts are rendered more salient.*

Note that this result is consistent with the self-concept maintenance theory proposed by Mazar *et al.* (2008). According to their framework, a moral prime renders non-criminal subjects more honest because it reminds them of their *own* moral standards. Our prime, however, was based on the immoral behaviour of a *third* person. One could even argue that when subjects are asked to think about the crimes of another person, their moral self-image is threatened and they, therefore, engage in counteractive behaviour (see Zhong and Liljenquist, 2006). This might explain why subjects in the criminal profile condition tended to be more honest than those in the control condition.

4. CONCLUSION

We study the impact of criminal identity salience on rule violating behaviour 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 coin tossing task. Our results show that prisoners cheated substantially more when their criminal identity was rendered more salient. This effect is specific to individuals with a criminal identity, as we did not find any influence of crime-related reminders in a placebo experiment with regular citizens. Furthermore, we show that our experimental measure of dishonesty correlates with inmates' disciplinary offences. Together, these findings suggest that criminal identity salience plays a crucial role in rule violating behaviour and also support recent theoretical endeavours incorporating identity into economic models of decision making (*e.g.* Akerlof and Kranton, 2000; Fang and Loury, 2005; and Bénabou and Tirole, 2011a).

It may seem surprising that we were able to increase the saliency of criminal identity of maximum security prisoners given that they are locked in an environment, which constantly reminds them of their criminal identities. However, our results from both the coin tossing task and the manipulation check suggest that everyday prison life does not necessarily render inmates' criminal identities maximally salient at all times. Inmates are probably able to distract themselves, and therefore habituate to the permanent cues in their environment.

From a policy perspective, our results raise the possibility that legal sanctions can have side effects with unintended consequences. Convictions often come along with collateral consequences, such as deprivations of civil rights, loss of professional licenses, or restricted access to public benefits (Travis, 2002). The intended deterrent effects of these consequences may come at the cost of reminding convicts of their criminal identity, which may increase their risk for relapse. As discussed above, however, permanent reminders are potentially less problematic than intermittent or novel cues. Identifying the extent to which the saliency of criminal identity

counteracts the deterrent effects of collateral consequences is, therefore, an important avenue for future research.

APPENDIX

A. Additional tables

TABLE A.1
Descriptive statistics of 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 offences	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

Notes: All variables are binary, except for sentence length (in years), years in prison, number of disciplinary offences, age (in years), risk attitudes (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 (χ^2 tests in case of binary variables and rank-sum tests in case of interval variables).

TABLE A.2
Descriptive statistics of 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 offences	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

Notes: All variables are binary, except for sentence length (in years), years in prison, and number of disciplinary offences. "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 (χ^2 tests in case of binary variables and rank-sum tests in case of interval variables).

TABLE A.3
Descriptive statistics of the placebo experiment

	Total sample $N=193$		Criminal identity $N=98$		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

Continued

TABLE A.3
Continued

	Total sample $N=193$		Criminal identity $N=98$		Control $N=95$		
Variable	mean	sd	mean	sd	mean	sd	p-value
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

Notes: All variables are binary, except for age (in years), risk attitudes (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 (χ^2 tests in case of binary variables and rank-sum tests in case of interval variables).

TABLE A.4
Descriptive statistics of total prison population based on the annual report

Variable	Total ($N=422$)	in %
Nationality:		
Swiss	111	26
South-eastern European	120	28
African	72	17
Central European	45	11
Other	74	18
Age (years):		
29 or younger	122	29
30–39	126	30
40–49	97	23
50–59	62	15
60 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

Acknowledgments. 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 Botond Kőszegi, two anonymous referees, Henk Aarts, John Antonakis, Dan Ariely, Daniel Benjamin, Sam Bowles, Alexander Cappelen, Daniel Chen, Stefano DellaVigna, Jérôme Endrass, Ernst Fehr, Nicola Gennaioli, 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 feedback. We acknowledge the support of the Gottlieb Duttweiler Institute.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

ABELER, J., BECKER, A., and FALK, A. (2014), "Representative Evidence on Lying Costs", *Journal of Public Economics*, **113**, 96–104.
ACEMOGLU, D. and JACKSON, M. O. (2014), "Social Norms and the Enforcement of Laws" (Working Paper No. 20369, NBER).
AKERLOF, G. A. and KRANTON, R. E. (2000), "Economics and Identity", *Quarterly Journal of Economics*, **115**, 715–753.
ALESINA, A., GIULIANO, P. and NUNN, N. (2013), "On the Origins of Gender Roles: Women and the Plough", *Quarterly Journal of Economics*, **128**, 469–530.
ANDERSON, D. A. (1999), "The Aggregate Burden of Crime", *Journal of Law and Economics*, **42**, 611–642.
ANDERSON, E. (2000), *Code of the Street: Decency, Violence, and the Moral Life of the Inner City* (New York: W.W. Norton & Company).
ARIELY, D. (2012), *The (Honest) Truth about Dishonesty: How we Lie to Everyone – Especially Ourselves* (New York: HarperCollins).
BALAFOUTAS, L., BECK, A., KERSCHBAMER, R. and SUTTER, M. (2013), "What Drives Taxi Drivers? A Field Experiment on Fraud in a Market for Credence Goods", *Review of Economic Studies*, **80**, 876–891.
BARGH, J. A. and CHARTRAND, T. L. (2000), "The Mind in the Middle: A Practical Guide to Priming and Automaticity Research", in Reis, H. T. and Judd, C. M. (eds.), *Handbook of Research Methods in Social Psychology* (New York: Cambridge University Press), 253–285.
BECKER, G. S. (1968), "Crime and Punishment: An Economic Approach", *Journal of Political Economy*, **76**, 169–217.
BEM, J. D. (1972), "Self-Perception Theory", in Berkowitz, L. (ed.), *Advances in Experimental Social Psychology*, Vol. 6 (New York: Academic Press), 1–62.
BÉNABOU, R. and TIROLE, J. (2011a), "Identity, Morals, and Taboos: Beliefs as Assets", *Quarterly Journal of Economics*, **126**, 805–855.
BÉNABOU, R. and TIROLE, J. (2011b), "Laws and Norms" (Working Paper No 15579, NBER).
BENJAMIN, D. J., CHOI, J. J. and FISHER, G. (2015), "Religious Identity and Economic Behavior" (Working Paper, Cornell University).
BENJAMIN, D. J., CHOI, J. J. and STRICKLAND, J. A. (2010), "Social Identity and Preferences", *American Economic Review*, **100**, 1913–1928.
BERTRAND, M., KAMENICA, E. and PAN, J. (forthcoming), "Gender Identitiy and Relative Income within Housholds" *Quarterly Journal of Economics*.
BIRKELAND, S., CAPPELEN, A. W., SØRENSEN, E. Ø. and TUNGODDEN, B. (2014), "An Experimental Study of Prosocial Motivation among Criminals", *Experimental Economics*, **17**, 501–511.
BRADLEY, M. M. and LANG, P. J. (1994), "Measuring Emotion: The Self-assessment Manikin and the Semantic Differential", *Journal of Behavior Therapy and Experimental Psychiatry*, **25**, 49–59.
BRADLEY, M. M. and LANG, P. J. (2000), "Measuring Emotion: Behavior, Feeling, and Physiology", in Lane, R., Nadel, L. and Ahern, G., (eds.), *Cognitive Neuroscience of Emotion* (Oxford University Press), 242–276.
CALLEN, M., ISAQZADEH, M., LONG, J. D. and SPRENGER, C. (2014), "Violent Trauma and Risk Preference: Experimental Evidence from Afghanistan", *American Economic Review*, **104**, 123–148.
CHALFIN, A. and McCRARY, J. (2014), "Criminal Deterrence: A Review of the Literature" (Working Paper, University of Cincinnati).
CHARNESS, G. and DUFWENBERG, M. (2006), "Promises and Partnership", *Econometrica*, **74**, 1579–1601.
CHARNESS, G., RIGOTTI, L. and RUSTICHINI, A. (2007), "Individual Behavior and Group Membership", *American Economic Review*, **97**, 1340–1352.
CHEN, Y. and LI, S. X. (2009), "Group Identity and Social Preferences", *American Economic Review*, **99**, 431–457.
CHEN, Y., LI, S. X., LIU, T. X. and SHIH, M. (2014), "Which Hat to Wear? Impact of Natural Identities on Coordination and Cooperation" *Games and Economic Behavior*, **84**, 58–86.
CHMURA, T., ENGEL, C. and ENGLERTH, M. (2013), "Selfishness as a Potential Cause of Crime. A Prison Experiment" (Working Paper, Max Planck Institute for Research on Collective Goods, Bonn).
COHEN, A. K. (1955). *Delinquent Boys: The Culture Of The Gang* (New York: The Free Press).
COHN, A., FEHR, E. and MARÉCHAL, M. A. (2014), "Business Culture and Dishonesty in the Banking Industry", *Nature*, **516**, 86–89.
DOHMEN, T., FALK, A., HUFFMAN, D., SUNDE, U., SCHUPP, J. and WAGNER, G. G. (2011), "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences", *Journal of the European Economic Association*, **9**, 522–550.
DRACA, M. and MACHIN, S. (2015), "Crime and Economic Incentives" (Working Paper, London School of Economics).

ELLINGSEN, T. and JOHANNESSON, M. (2004), "Promises, Threats and Fairness", *Economic Journal*, **114**, 397–420.
FANG, H. and LOURY, G. C. (2005), "Dysfunctional Identities can be Rational", *American Economic Review*, **95**, 104–111.
FISCHBACHER, U. and FÖLLMI-HEUSI, F. (2013), "Lies in Disguise - An Experimental Study on Cheating", *Journal of the European Economic Association*, **11**, 525–547.
FREDERICK, S. (2005), "Cognitive Reflection and Decision Making", *Journal of Economic Perspectives*, **19**, 25–42.
GENDREAU, P., LITTLE, T. and GOGGIN, C. (1996), "Meta-Analysis of the Predictors of Adult Offender Recidivism: What Works!", *Criminology*, **34**, 575–607.
GNEEZY, U. (2005), "Deception: The Role of Consequences", *American Economic Review*, **95**, 384–394.
GOETTE, L., HUFFMAN, D. and MEIER, S. (2012), "The Impact of Social Ties on Group Interactions: Evidence from Minimal Groups and Randomly Assigned Real Groups", *American Economic Journal: Microeconomics*, **4**, 101–115.
HOFF, K. and PANDEY, P. (2006), "Discrimination, Social Identity, and Durable Inequalities", *American Economic Review*, **96**, 206–211.
HOFF, K. and PANDEY, P. (2014), "Making up People - the Effect of Identity on Performance in a Modernizing Society", *Journal of Development Economics*, **106**, 118–131.
HOUSER, D., VETTER, S. and WINTER, J. (2012), "Fairness and Cheating", *European Economic Review*, **56**, 1645–1655.
HOWARD, B. S. (1963), *Outsiders: Studie in the Sociology of Deviance* (New York: The Free Press).
KAPLAN, H. B. (1975). *Self-Attitudes and Deviant Behavior* (Oxford: Goodyear).
KARTIK, N. (2009), "Strategic Communication with Lying Costs", *Review of Economic Studies*, **76**, 1359–1395.
KESSLER, J. B. and MILKMAN, K. L. (2014), "Identity in Charitable Giving" (Working Paper, The Wharton School, University of Pennsylvania).
KHADJAVI, M. and LANGE, A. (2013), "Prisoners and their Dilemma", *Journal of Economic Behavior & Organization*, **92**, 163–175.
KRANTON, R. E., PEASE, M., SANDERS, S. and HUETTEL, S. (2013), "Identity, Groups, and Social Preferences" (Working Paper, Duke University).
LeBOEUF, R. A., SHAFIR, E. and BAYUK, J. B. (2010), "The Conflicting Choices of Alternating Selves", *Organizational Behavior and Human Decision Processes*, **111**, 48–61.
LERMAN, A. E. (2009), "The People Prisons Make: Effects of Incarceration on Criminal Psychology", in Steven, R. and Stoll, M. (eds.), *Do Prisons Make Us Safer?: The Benefits and Costs of the Prison Boom* (New York: Russell Sage Foundation), 151–176.
LUDWIG, J. (2006). *Testimony to the U.S. Senate Judiciary Committee*, 19 September 2006. retrieved 30 April 2015 http://gpo.gov/fdsys/pkg/CHRG-109shrg42938/html/CHRG-109shrg42938.htm.
MAZAR, N., AMIR, O. and ARIELY, D. (2008), "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance", *Journal of Marketing Research*, **45**, 633–644.
McADAMS, R. H. and RASMUSEN, E. B. (2007), "Norms and the Law", in Polinsky, M. A. and Shavell, S. (eds.), *Handbook of Law and Economics*, Vol. 2 (Amsterdam: Elsevier), 1573–1618.
NAGIN, D. S. (2013), "Deterrence: A Review of the Evidence by a Criminologist for Economists", *Annual Review of Economics*, **5**, 83–105.
PEW CENTER ON THE STATES (2011). *State of Recidivism: The Revolving Door of America's Prisons* (Washington, DC: Pew Charitable Trust).
POSNER, E. A. (2009), *Law and Social Norms* (Cambridge: Harvard University Press).
PRUCKNER, G. J. and SAUSGRUBER, R. (2013), "Honesty on the Streets: A Field Study on Newspaper Purchasing", *Journal of the European Economic Association*, **11**, 661–679.
RECKLESS, W. C. (1967), *The Crime Problem* (New York: Appleton-Century-Crofts).
SHALVI, S., DANA, J., HANDGRAAF, M. J. J. and DeDREU, C. (2011), "Justified Ethicality: Observing Desired Counterfactuals Modifies Ethical Perceptio ns and Behavior", *Organizational Behavior and Human Decision Processes*, **115**, 181–190.
SHERMAN, L. W. (1993), "Defiance, Deterrence, and Irrelevance: A Theory of the Criminal Sanction", *Journal of Research in Crime and Delinquency*, **30**, 445–473.
SHIH, M., PITTINSKY, T. L. and AMBADY, N. (1999), "Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance", *Psychological Science*, **10**, 80–83.
SIMOURD, D. J. and OLVER, M. E. (2002), "The Future of Criminal Attitudes Research and Practice", *Criminal Justice and Behavior*, **29**, 427–446.
TRAVIS, J. (2002), "Invisible Punishment: An Instrument of Social Exclusion", in Chesney-Lind, M. and Mauer, M. (eds.), *Invisible Punishment: The Collateral Consequences of Mass Imprisonment* (New York: The New Press).
WALTERS, G. D. (2003), "Changes in Criminal Thinking and Identity in Novice and Experienced Inmates Prisonization Revisited", *Criminal Justice and Behavior*, **30**, 399–421.
ZHONG, C.-B. and LILJENQUIST, K. (2006), "Washing Away Your Sins: Threatened Morality and Physical Cleansing", *Science*, **313**, 1451–1452.