ELSEVIER

Evolution and Human Behavior xx (2012) xxx – xxx

Evolution and Human Behavior

Original Article

To punish or repair? Evolutionary psychology and lay intuitions about modern criminal justice

Michael Bang Petersen[a,*], Aaron Sell[b,c], John Tooby[b], Leda Cosmides[b]

[a]Department of Political Science and Government, Aarhus University, Aarhus, Denmark
[b]Center for Evolutionary Psychology, University of California, Santa Barbara, CA, USA
[c]School of Criminology and Criminal Justice, Griffith University, Australia

Initial receipt 4 January 2011; final revision received 7 May 2012

Abstract

We propose that intuitions about modern mass-level criminal justice emerge from evolved mechanisms designed to operate in ancestral small-scale societies. By hypothesis, individuals confronted with a crime compute two distinct psychological magnitudes: one that reflects the crime's seriousness and another that reflects the criminal's long-term value as an associate. These magnitudes are computed based on different sets of cues and are fed into motivational mechanisms regulating different aspects of sanctioning. The seriousness variable regulates how much to react (e.g., how severely we want to punish); the variable indexing the criminal's association value regulates the more fundamental decision of how to react (i.e., whether we want to punish or repair). Using experimental designs embedded in surveys, we validate this theory across several types of crime and two countries. The evidence augments past research and suggests that the human mind contains dedicated psychological mechanisms for restoring social relationships following acts of exploitation.

Keywords: Criminal justice; Punishment; Reparation; Recalibrational theory; Evolutionary psychology

1. Introduction

As in other species, the social world of our ancestors contained individuals who were poised to exploit others if such acts were self-beneficial (Daly & Wilson 1988; Duntley & Buss 2004; Duntley 2005). This selection pressure favored the evolution of psychological mechanisms designed to counter exploitation of one's self, family, and social group through punishment (Boehm 1985; Daly & Wilson 1988; Frank 1988; Duntley & Buss 2004; Sell, Tooby, & Cosmides 2009; Petersen, Sell, Tooby, & Cosmides 2010; McCullough, Kurzban, & Tabak 2011). Modern crimes have features that satisfy the input conditions of mechanisms designed to respond to exploitation, and recent research suggests that our evolved counterexploitation psychology structures the intuitions that modern individuals have about criminal justice (Aharoni & Fridlund, 2011; Petersen et al. 2010; Robinson, Kurzban, & Jones 2007). This research has documented high levels of cross-cultural agreement concerning the seriousness of different crimes (Robinson et al. 2007). Furthermore, with considerable supporting evidence, it has been argued that this perceived seriousness taps into our evolved sense of justice, such that individuals prefer sanctions that are proportional to the seriousness of the crime (e.g., Darley & Pittman 2003; Aharoni & Fridlund 2011).

Although this research has provided important insights, both the evolutionary literature on exploitation and its applications to modern criminal justice have neglected the existence of counterexploitation strategies beyond punishment. The small-scale social world of our ancestors—with dense social networks and high levels of dependency—should have selected for nonpunitive reparative strategies, in addition to punitive ones (Aureli & de Waal 2000; Petersen et al. 2010; McCullough et al. 2011). We propose that a major factor regulating the activation of reparative—rather than punitive—responses to rule violations is the (perceived) social value of the perpetrator. If so, then our intuition that more serious crimes call for more serious sanctions should be joined by another intuition: that different types of sanction are appropriate, depending on perceptions of the criminal's social worth.

* Corresponding author. Tel.: +45 89 42 54 26.
E-mail address: michael@ps.au.dk (M.B. Petersen).

1090-5138/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.evolhumbehav.2012.05.003

An evolutionary and computational dissection of exploitation led us to predict that the human mind spontaneously computes the magnitudes of two distinct psychological variables when confronted with exploitation. One represents the exploitation's seriousness, as stressed by previous theories. The other represents the exploiter's association value—the person's value as a potential associate (Petersen et al. 2010). By hypothesis, the magnitude of each variable is computed based on different sets of cues, and these variables are fed into motivational mechanisms regulating distinct aspects of strategies for countering exploitation. Whereas the indexed seriousness of an exploitive act regulates how much to react (e.g., how severely we want to punish, how long we may wish to incapacitate the perpetrator, or how intense our efforts at social repair will have to be), the exploiter's indexed social or association value regulates the more fundamental decision of how to react (i.e., whether we want to punish or repair). We have termed this theory the recalibrational theory of counterexploitation (Petersen et al. 2010). In this article, we provide empirical evidence for this theory from the intuitions of lay individuals about criminal justice. We argue that current models of criminal justice intuitions should be expanded to account for the existence and effects of nonpunitive reparative sentiments in the human response to exploitation and crime (see also Aureli & de Waal 2000; McCullough et al. 2011).

2. The recalibrational theory of counterexploitation

In the small-scale settings of our ancestors, actions would often have had consequences for individuals beyond the actor. When this holds true, mechanisms in the mind must decide how much to weigh the other person's welfare relative to the actor's own. Recent research demonstrates that social decisions depend upon the magnitude of an internal variable—a welfare tradeoff ratio (WTR)—which sets the weight the actor places on a specific person's welfare relative to the actor's own (Tooby, Cosmides, & Price 2006; Tooby & Cosmides 2008; Sell et al. 2009; Delton 2010). The higher an actor's WTR toward the target person, the more the actor will sacrifice his or her welfare to enhance the target's welfare. The lower an actor's WTR toward a target, the more likely the actor is to harm the target when doing so is personally beneficial.

Within this framework, we can define exploitation as acts expressing too low a WTR (relative to some baseline) by inflicting a cost on the target for too small a benefit to oneself. Given the acuteness of this adaptive problem, evolution should have selected for counterexploitation strategies that are designed to recalibrate the exploiter's WTRs because increasing the magnitude of these variables should decrease the number of exploitive acts that they commit in the future (see also Sell et al. 2009). On this view, some counterexploitation strategies—including punishment—are recalibrational strategies.

By changing the costs of exploitive acts, punishment serves a recalibrative function: it induces the exploiter to place greater weight on the welfare of others in the future (Jacoby 1983; Daly & Wilson 1988; Clutton-Brock & Parker 1995; de Waal 1996; Fehr & Gächter 2002; Fehr & Fischbacher 2004). However, punishment has a shortcoming: its efficacy as a counterexploitation strategy is fully contingent on the punisher's ability to monitor the exploiter's behavior. For example, experimental evidence from economic games demonstrates that punishment powerfully reduces free riding. If the possibility of punishment is removed in later rounds, however, free riding again rapidly increases (e.g., Fehr & Gächter 2000, 2002). Hence, punishment works, but only within certain limits. Because much behavior is not monitored by others, reliance on punitive strategies leaves exploitation uncountered across a broad range of conditions.

We argue that reparative strategies were, in part, selected to remedy this problem (de Waal 1996; Petersen et al. 2010; McCullough et al. 2011). Anthropologists have documented restorative sanctions across diverse small-scale and agricultural societies (Fry 2000; Braithwaite 2002). Similar, if sparser, observations have been made by primatologists who interpret certain behaviors as reconciliatory acts used to manage conflicts and aggressive encounters (de Waal 1996; Aureli & de Waal 2000). Research on reparative gestures demonstrates that they involve demonstrations of how the exploiter's behavior violated social obligations (Vangelisti, Daly, & Rudnick 1991), reminding the exploiter of favors done for them in the past (Sell 2005) or signaling a wish for future prosocial interaction (Fujisawa, Kutsukake, & Hasegawa 2005). Similarly, the reconciliation rituals of nonhuman primates involve grooming—a benefit normally exchanged among social partners (de Waal 1996).

These reparative gestures convey information to exploitive persons that they have underestimated the true magnitude of the harm inflicted, underestimated the true value of the relationships jeopardized, or overestimated the gain to the exploiter of acting selfishly when compared to the magnitude of the loss inflicted on the other party. Such information, we argue, targets WTR circuits that are distinct from those targeted by punishment. Because different factors regulate whether a specific action is adaptive in private vs. public contexts, evolution should have selected for machinery designed to compute a monitored WTR to govern decisions when one's actions are likely to become known to those who will be affected by them, and a different, intrinsic WTR to govern decisions when one's actions are not being monitored. Whereas monitored WTRs are expected to be influenced by the ability of the target to respond to actions affecting his or her welfare, for example, by inflicting costs in the form of punishment, such factors are less relevant in setting intrinsic WTRs. Here, other factors should emerge as important regulators of welfare tradeoffs, such as whether the actor and target are kin (Lieberman & Linke 2007) or whether the actor's welfare is yoked to that of the target

(Tooby & Cosmides 1996). By conveying information about the value to the exploiter of his relationship to those harmed, reparative gestures aim at up-regulating the exploiter's intrinsic WTRs, such that they inflict less harm in the future even when not being monitored. In line with this, research on emotions suggests that such reparative interventions, when successful, elicit guilt in exploiters (Harris, Walgrave, & Braithwaite 2004), which subsequently up-regulates their cooperativeness (Harris 2003; Ketelaar & Au 2003). Furthermore, research has produced evidence consistent with the notion that feelings of guilt correspond specifically to an up-regulation of the exploiter's intrinsic rather than monitored WTR (cf. Tooby & Cosmides 2008): reparative feelings of guilt translate into motivations to spontaneously help, donate, and share, but these intrinsic prosocial motivations are reduced by fear of punishment (Caprara, Barbaranelli, Pastorelli, Cermak, & Rosza 2001).

2.1. Deciding the intensity and type of sanction

To the extent evolution has selected for both punitive and nonpunitive strategies, the human mind is required to solve two distinct computational problems when confronted with exploitive acts. First, as punitive and reparative goals are often incompatible, the mind needs to solve the qualitative problem of choosing between a punitive or rehabilitative reaction. Second, the mind then needs to solve the quantitative problem of measuring out the adaptive level of cost-imposition or guilt-induction, that is, to determine how much the targeted WTR needs to be recalibrated.

2.1.1. Intensity of reaction

The need for a strict quantitative modulation of recalibrational strategies arises from the fact that both punitive and reparative gestures should be allocated with care. Punishment is costly and can also trigger retaliation. Similarly, excessive demands for remorse in reparative strategies can backfire and further unravel social relationships. Within the recalibrational framework, the problem posed by exploitation is that it indicates that the perpetrator does not value the victim or other members of one's social group sufficiently, predicting future exploitation. Together with the benefit of the act to the perpetrator, the costliness of the act to the victim provides a reliable indication of the offender's current WTR toward the target: that is, how little the offender values the welfare of the target compared to his own.

We propose that the difference between the minimally acceptable WTR and the WTR expressed toward the victim reflects the intuitive concept of an exploitive act's seriousness. In the recalibrational theory, the seriousness of an offense thus tells an individual how much the perpetrator's WTR toward potential targets needs to be up-regulated before being deemed sufficiently high. This information should be reflected in the intensity of the reaction by which we seek to achieve this objective, whether punitive or rehabilitative.

In line with this evolutionary and computational account of the concept of seriousness, a large criminological literature has documented that (i) the seriousness of a crime is set by the crime's physical or symbolic level of harm, (ii) citizens are easily able to rank different crimes with respect to their seriousness, and (iii) these rank orderings are stable across cultures (at least with respect to crimes causing physical harm) (Stylianou 2003). Although less explored, there is also evidence that the same crime (i.e., same cost inflicted) is seen as less serious when done for a large gain in inclusive fitness, such as to feed one's family, than for a small gain (Rossi, Simpson, & Miller 1985). This suggests that intuitions about crime seriousness track welfare trade-offs rather than harm alone (see also Duntley & Buss 2004).

2.1.2. To punish or repair?

The choice between reparative and punitive strategies should reflect the merits of targeting intrinsic and monitored WTRs, respectively. If it is possible to up-regulate the target's intrinsic WTR toward oneself and those one values, then the target will spontaneously consider your welfare and theirs, even when you are absent or temporarily incapacitated. This is an efficient outcome: one harvests the benefits of associating with the target without having to constantly monitor the target's behavior to prevent opportunistic acts of exploitation. The intuition that the offender can be successfully recalibrated is reflected in the colloquial expression, "he can be made to realize his mistake so that he does not repeat it."

Still, it may be difficult to up-regulate an exploiter's intrinsic WTRs broadly, toward all potential victims. Conversely, punitive strategies involve actual changes to the cost–benefit ratios associated with exploitive acts and, hence, can revise monitored WTRs in a way that protects broad classes of potential victims. The up-regulation of monitored WTRs may be more effective at preventing harm; however, they are less likely to elicit a flow of resources or other benefits from a target whose intrinsic WTRs are too low.

The activation of punitive and reparative strategies requires the individual to gauge and weigh the potential benefits and risks of associating with the exploiter in the future (see also de Waal 1996; McCullough et al. 2011). The seriousness of an exploitive act does provide some information about these benefits and risks—a particularly heinous crime suggests that the criminal's WTRs are extremely low, making him dangerous to be around and perhaps difficult to recalibrate. Greater seriousness should therefore, ceteris paribus, increase preferences for punitive over reparative strategies. All else is never equal, however, and other factors can increase the expected payoff of a reparative counterstrategy.

The value of a reparative strategy is higher if it is likely to result in the successful recalibration of the exploiter and if

the reformed exploiter has the potential to contribute to the welfare of others. The exploiter showing remorse or having a stake in one's welfare by virtue of kinship or friendship (e.g., Lieberman & Linke 2007) are factors suggesting that it may be possible to raise the exploiter's intrinsic WTRs toward others; his being an in-group member or high in status suggests that attempts to reform the offender will benefit others if they succeed. All of these factors can contribute to the judgment that, under the right circumstances, the offender might become a valuable associate in the future (see Fig. 1).

The human evolved psychological architecture should contain subcomponents designed to track cues to these factors and use them to compute mental representations of the association value of others—that is, the estimated net lifetime future value of maintaining interactions with another from the point of view of the decision maker (Tooby & Cosmides 1996). Fig. 1 provides an overview of the association value computational system. As seen in the left side of the figure, many factors can render a person valuable as an associate, such as their status, willingness to cooperate, and the fact that they are kin. Some of these factors raise Y's value as an associate because they predict that Y's intrinsic WTR toward X will be higher than that of many others (e.g., Y is X's brother, friend, or mate). A higher intrinsic WTR means, ceteris paribus, that Y will sacrifice more to provide benefits to X and engage in fewer acts that exploit X (foregoing more benefits to avoid imposing costs on X). Y's association value can also be raised by factors not directly linked to Y's intrinsic WTRs. Y may be very skilled at procuring resources (e.g., Y is a good hunter), in a position to confer benefits (e.g., Y has high status), or likely to emit positive externalities (e.g., Y is a formidable in-group member capable of defending X's group from attack). All of these factors should affect the association value that X assigns to Y (and this association value should, in turn, be one of the factors affecting X's monitored and intrinsic WTRs toward Y). Behaviorally, a high association value predicts two classes of outcome: (i) Y will be less willing to impose costs on X and (ii) Y will be more willing or able to provide benefits to X. This should be reflected in X's expectations about Y's future behavior, as shown in the right side of the figure.

We expect counterexploitation circuitry to access the index of the exploiter's association value index and use its magnitude to regulate the activation of punitive and reparative strategies. When an exploiter's association value is indexed as high, the sentiment that reparative strategies are appropriate will be triggered; when low, punitive strategies will be activated. This latter expectation is consistent with a diverse range of evidence. Fry (2000) sums up the anthropological observations on reconciliation by arguing that, cross-culturally, it seems most likely to occur when relationships are important and difficult to replace. Furthermore, evidence from studies of children's conciliatory behavior suggests that the frequency of reconciliation is higher among friends than acquaintances (Cords & Killen 1998; Fujisawa et al. 2005). Finally, both experimental work

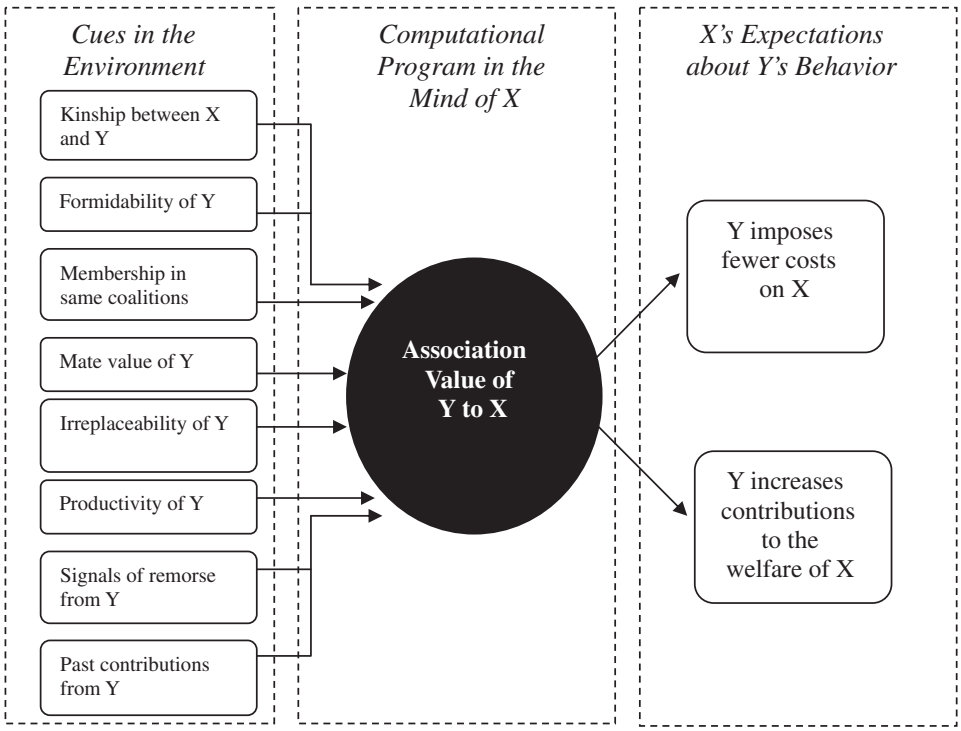

Fig. 1. The causes and behavioral effects of association values.

and observations in primate groups suggest that the likelihood of reconciliation following hostile encounters is correlated with the social value of the relationship (Aureli & de Waal 2000; although the primate data must be viewed cautiously; see Silk 2002).

3. Predictions: the recalibrational theory and modern criminal justice

Modern crimes such as vandalism, theft, violence, and murder involve the imposition of costs for personal benefit and therefore have features that should trigger evolved intuitions about countering exploitation. If so, the recalibrational theory should allow us to predict under which circumstances lay people prefer different types and intensities of criminal justice sanctions. Yet, whereas modern society is large scale, ancestral social interaction was small scale (Kelly 1995). How modern individuals reason about criminal justice can, in other words, be expected to be influenced by the factors that were important in ancestral small-scale counterexploitation, even if they no longer make sense (Petersen, 2012).

In large-scale modern societies, criminal justice can assume two forms: punishment and/or rehabilitation. The activation of punitive motivations should facilitate support for prison time, fines, and other modern forms of punishment, whereas the triggering of reparative motivations should facilitate support for modern forms of rehabilitation. According to Cullen and Gendreau (2000; p. 112), rehabilitation is an intentional intervention to change a rule violator and prevent new violations without installing fear of sanctions. Rehabilitation often involves counseling, drug treatment, and other interventions designed to help the offender realize the harm they have done and increase the extent to which the offender values the welfare of others.

Applying the recalibrational theory of counterexploitation to modern criminal justice intuitions, we can outline three predictions. To the extent that our argument is correct, the variable reflecting an exploiter's association value and the variable that reflects the seriousness of the exploitive act have highly distinct regulatory functions within the brain's counterexploitation system. Furthermore, these two magnitudes are expected to be represented as distinct from each other and to be computed from different sets of cues:

Prediction 1. The perceived association value of the criminal should predict whether a punitive or reparative sanction is preferred, while the seriousness of the crime should be a much smaller predictor of this preference.

Prediction 2. The seriousness of a crime should predict the preferred intensity of the sanction, while the perceived association value of the criminal should be a much smaller predictor of sanction intensity.

Prediction 3. Ecologically valid cues of the probability that the criminal will impose costs or contribute benefits in the future should be spontaneously incorporated into the computation of the criminals' association value but only marginally into the perceived seriousness of the crime.

3.1. Testing against alternative hypotheses

The best-known and supported theory of criminal justice intuitions is just deserts theory. According to this theory, people prefer sanctions that are proportional to the seriousness of the crime, where seriousness is primarily (although not exclusively) conceptualized as the extent of inflicted harm (Warr, Meier, & Erickson 1983; Darley, Carlsmith, & Robinson 2000; Darley & Pittman 2003; Roberts & Stalans 2004). Prediction 2 of the recalibrational theory of counterexploitation overlaps with just deserts theory, but Predictions 1 and 3 are new.

Previous research has consistently shown that sanctioning preferences in specific crime cases are highly correlated with the perceived seriousness of the crime, but not with perceptions of the criminal's future behavior (Warr et al. 1983; Darley et al. 2000; Carlsmith, Darley, & Robinson 2002). These observations invited the conclusion that criminal justice intuitions are driven by a retrospective desire for revenge; accordingly, the theory holds that sanctioning preferences are determined exclusively by circumstances in the past, which shape people's perceptions of how serious the crime was (Darley & Pittman 2003). If it were true that expectations about future behavior—whether exploitive (as in recidivism) or productive—play no role in determining what kind of sanctions laymen prefer, that would be evidence against the recalibrational theory tested herein. However, the past studies that appear to support this claim did not consider the role of reparative sentiments in depth and focused almost exclusively on punishment severity.

According to the recalibrational theory of counterexploitation, the severity of punishment should indeed be modulated be the seriousness of the crime (cf. Prediction 2). More serious crimes indicate a lower WTR; the lower the exploiter's WTR, the more costs must be inflicted on the exploiter to motivate an up-regulation of sufficient magnitude in his WTR. The same logic applies to reparative strategies: more serious crimes indicate lower WTRs, suggesting that more intense reparative efforts are required to up-regulate intrinsic WTRs to an acceptable level.

The predictions derived from recalibrational theory diverge from just deserts theory when one considers not the severity of punishment but the more fundamental choice of whether to punish or rehabilitate. For the choice to punish or repair, perceptions of the criminal—including his or her future behavior—should be at least as important as perceptions of the crime. In fact, displays of remorse, lack of intent, and lack of a criminal record are often considered mitigating circumstances, leading people to prefer less severe sanctions. To explain this, just deserts theory claims that these factors lower the perceptions of how serious the

crime was (a proposition we test herein) (Carlsmith et al. 2002; Roberts 2008). Conceptually, this is plausible for the lack of intent (to harm), but it is not at all clear why a crime should be seen as less serious—as having caused less harm—because the criminal later repents or because it was his first offense. These factors seem to say more about the criminal than the crime, which fits more comfortably with the recalibrational theory, as outlined below.

4. The present studies

In line with previous research on the psychology of criminal justice, we test our predictions by analyzing subjects' responses to vignettes about crimes (Study 1: vandalism, robbery with assault, rape; Study 2: battery). Two studies were conducted, one in Denmark and the other in the United States, to provide cross-cultural leverage. The American and Danish criminal justice systems are very different, as are the attitudes of their general populations toward criminal justice (Selke 1991). Whereas the Danish system and the Danes have considerable focus on and support for rehabilitation, the United States has the highest prison incarceration rates in the world, and Americans are highly supportive of punishment. Consistent with an evolutionary psychology perspective, a successful replication of the predicted effects between these two countries would suggest that they do not reflect feedback from the workings of a particular criminal justice system.

While previous accounts and the recalibrational theory all stress the importance of inflicted costs for individuals' computations of crime seriousness, the recalibrational theory suggests that individuals also compute a psychologically distinct decision variable: the criminal's association value. This computation should be based on certain cues that are distinct from those by which seriousness is gauged, such as whether the criminal is an in-group or out-group member. In this way, the recalibrational theory provides a new understanding of why many mitigating circumstances matter: lack of a criminal record, lack of intent, and remorse are cues to the future value of the criminal rather than to the seriousness of his crime, as argued in extant research (cf. Carlsmith et al. 2002; p. 285). History predicts future behavior; lack of intentionality suggests that the criminal's actions did not result from a low WTR toward the victim; remorse indicates that the criminal realizes that he placed insufficient weight on the welfare of the victim and plans to recalibrate his WTR upward so that he does not do so in the future.

In our theory, many of the factors thought to be influencing public perceptions of what the criminal has done instead influence perceptions of what the criminal will do in the future and, therefore, his value as an associate. In our tests of the recalibrational theory, we focus on these novel predictions regarding the role played by perceived association values in criminal justice intuitions. In each study, we experimentally varied mitigating factors that were thought to influence perceptions of crime seriousness. If the recalibrational theory is correct, these factors should regulate the choice between punitive and reparative preferences, not by influencing how serious the crime is perceived to be but by influencing the perceived association value of the criminal.

As described above (see Fig. 1), the representation that an individual has a high association value is tied to two expectations concerning the future behavior of that individual: that the individual is likely to (i) avoid imposing costs and/or (ii) provide benefits. Given the existence of these two theoretically valid expressions of a high association value index, we use both—one for each of the two studies.

In Study 1, perceived association value was operationalized by a measure that captures the criminal's probability of deciding to impose costs by committing criminal acts in the future (i.e., recidivism). The inference that a person has a high probability of exploiting others should, ceteris paribus, reduce that person's association value. If criminal behavior is a form of exploitation, then the perception that an individual is likely to commit criminal acts in the future should drastically lower their value as an associate.

In Study 2, perceived association value was operationalized by a measure that captures the criminal's probability of becoming a productive member of society in the future. Again, ceteris paribus, the inference that a person has a low probability of becoming a productive group member reduces that person's association value.

5. Study 1

5.1. Methods

In Study 1, we analyzed subjects' responses to three vignettes about crimes that varied in seriousness (vandalism, robbery involving assault, rape). These were embedded in a paper-and-pencil survey given to 4116 Danish upper-secondary students aged 15–21 years, representing a broad cross section of Danish society (see below for descriptions and online supplemental materials for full wordings, available on the journal's website at www.ehbonline.org). The second and third vignettes contained experiments specifically designed to manipulate the association value of the criminal.[1]

To test the recalibrational theory, we obtained the subjects' perception of the association value of each criminal. Following the logic that higher association values generate the expectation that the person is less likely to impose costs in the future, the association value measure in Study 1 tapped the subject's judgment of how likely the criminal would be to commit similar acts in the future.

[1] The first vignette also contained an experiment. This experiment was not designed to discriminate between association values and crime seriousness, and hence, we do not analyze the experimental manipulation in the main text. See the online supplemental materials (available on the journal's website at www.ehbonline.org) for analyses.

Subjects were also asked to rate the seriousness of the crime, whether they preferred a reparative or punitive sanction, and the preferred intensity of the sanction. The exact wording of all of the survey items appears in the online supplemental materials (available on the journal's website at www.ehbonline.org).

In all cases, the variables were coded between 0 and 1 (high values indicating high perceived association value, high seriousness, a preference for an intense sanction, and support for reparative over punitive sanctions, respectively). All analyses are based on linear regression with unstandardized regression coefficients used as effect sizes (see online supplemental materials for details, available on the journal's website at www.ehbonline.org). Because we are interested in whether perceptions of association value and seriousness are regulated by different cues and regulate distinct aspects of counterexploitation strategies, the analyses compute, first, their independent effects on outcome variables, controlling for one another and all other measures and, second, the partial effect of cues on perceptions (Vignettes 2 and 3), controlling for the other perception. This provides the strongest test of the hypotheses.

5.2. Results

5.2.1. Vignette 1: robbery

The first vignette presents a case of assault and robbery, a crime of medium seriousness. Analyzing subjects' responses to the vignette allows us to test Predictions 1 and 2.

Does perceived association value predict preferences for reparative over punitive sanctions? Yes. The preference for reparative over punitive strategies was correlated with the perceived association value of the criminal (b=0.38, p<.001). As predicted, higher association value—operationalized as lower probability of committing similar acts in the future—was a robust and independent predictor of the extent to which subjects preferred reparative over punitive strategies, but the perceived seriousness of the crime was not (b=−0.04, p=.185).

Does the perceived seriousness of the crime predict the preferred intensity of the sanction? Yes. As in previous research, subjects who perceived the crime as more serious preferred more intense sanctions (b=0.39, p<.001). In contrast to past studies claiming that the intensity of the sanction is unaffected by expectations concerning the criminal's future behavior, there was also an independent effect of the criminal's association value, meaning that less intense sanctions were preferred by those who thought the criminal was unlikely to repeat his actions in the future. However, the effect size (b=−0.19; p<.001) was remarkably small, given that it would be reasonable to assume that more intense sanctions are necessary to deter a person who is otherwise likely to repeat his crime. Consistent with Prediction 2, the effect size for perceived seriousness was two times larger than for association value.

Unlike other theories, the recalibrational theory predicts that people's perceptions of the association value of criminals will regulate their preferences for reparative over punitive sanctions. This novel prediction was supported by the results of the first vignette. The results also supported the claim that the crime's seriousness will regulate the intensity of the sanction, as predicted by both the recalibrational theory and just deserts theory. In contrast to just deserts (but consistent with the recalibrational theory), association value also had a slight impact, such that those who thought the criminal was unlikely to repeat his crime in the future favored less intense sanctions.

5.2.2. Vignette 2: vandalism

The second vignette was designed to (i) test the robustness of our findings, (ii) generalize them to the class of nonviolent crimes, and (iii) test whether ecologically valid cues of future value contribute to perceptions of association value but not the crime's seriousness (Prediction 3). The vignette focuses on a case of vandalism and includes a between-subjects experiment, which manipulates the past behavior of the criminal. In one condition, subjects are informed that the criminal has three similar crimes from earlier on his record. In the other condition, subjects are informed that the criminal has no prior criminal record.

Past criminal behavior was manipulated because it is an ecologically valid cue to future criminal behavior, which should be a major factor determining the criminal's future value as an associate. Unsurprisingly, the manipulation was successful: ratings of association value were higher for the vandal with no criminal record than for the one with a history of vandalism (M=0.81 vs. 0.23, p<.001). By hypothesis, preference for reparative over punitive strategies is most strongly regulated by association value, so association value should mediate any relationship that might exist between past criminal behavior and reparative preferences. The results of this mediation analysis using linear regression models are presented in Fig. 2.

Does perceived association value predict preferences for reparative over punitive sanctions? Yes. The preference for reparative over punitive strategies was correlated with the perceived association value of the criminal (b=0.27), replicating the relationship found for the robbery vignette. Crime seriousness also had a significant effect (with less serious crimes predicting preferences for more reparative sanctions), but it was small (b=−0.14)—half the size of the effect for association value. A formal test of their absolute values demonstrates that the effect size for association value is significantly higher than the effect size for crime seriousness (F=10.80, p=.001).

Does an ecologically valid cue of association value—the criminal's past behavior—contribute more strongly to perceptions of association value than to perceptions of the seriousness of his crime? Yes. Variation in information about the criminal's past behavior had a large effect on perceptions

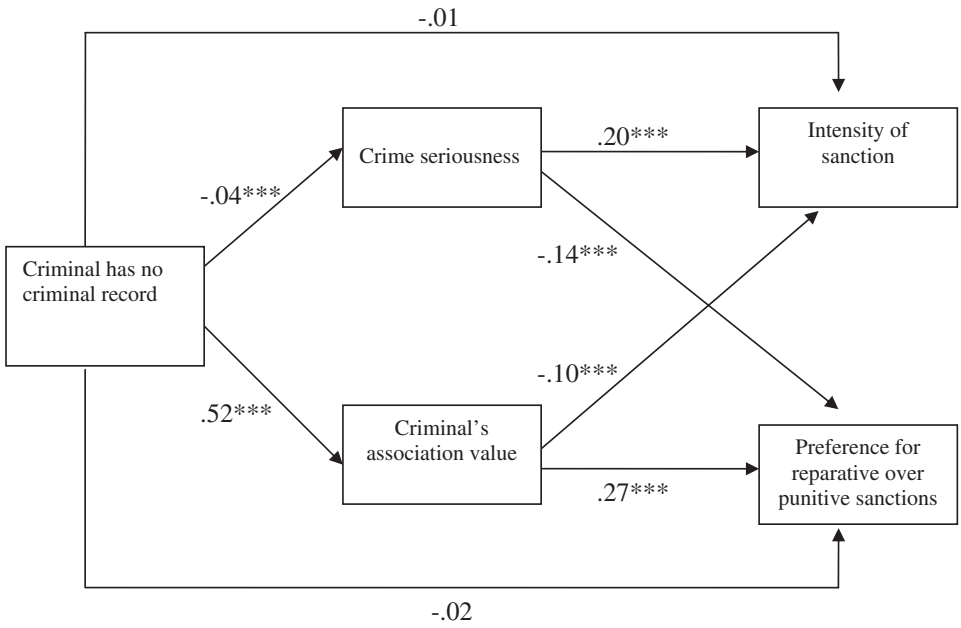

Fig. 2. Vignette 2, Study 1 (vandalism): perceived seriousness of crime and perceived association value of the criminal as mediators of a specific criminal's past behavior. Unstandardized regression coefficients (*b*): all variables are coded between 0 and 1.

of his association value (b=0.52), but very little effect on how serious his crime was seen to be (b=−0.04). The effect of history on association value was 13 times greater than its effect on perceptions of the crime's seriousness.

Does the criminal's past behavior have a direct effect on reparative sentiments? No. Information about the criminal's past behavior affects perceptions of his future association value and the seriousness of his crime, and these perceptions fully mediate the relationship between criminal history and preference for reparative over punitive sanctions. After controlling for these perceptions, the relationship between criminal history and reparative preferences is no longer significant (before control: b=0.17, p<.001; after control: b=−0.02, p=.18). Sobel tests of mediation reveal two significant paths: a strong path through perceived association value (effect size=0.14, z=13.07, p<.001) and a weak path through perceived seriousness (effect size=0.01, z=3.04, p=.002). The path through perceived association value is so strong that controlling for that variable is, by itself, sufficient to render the relationship between criminal history and reparative preferences insignificant (after control: b=−0.02, p=.31).

Does the perceived seriousness of the crime predict the preferred intensity of the sanction? Yes. Preferred intensity of sanction was predicted by both the perceived seriousness of the crime (b=0.20, p<.001) and perceptions of the criminal's association value (b=−0.10, p<.001), but the effect size for seriousness was twice as great as that for association value. This difference between the absolute effect sizes is highly significant (F=31.33, p<.001).

Taken together, these results support the claim that the association value of a criminal is psychologically distinct from perceptions of the crime's seriousness, both in how it is computed and the effects it has on other judgments. Criminal history is weighted far more heavily by the system that computes association value than by the system computing the seriousness of the crime: information about the vandal's past criminal activities had a large effect on perceptions of his association value, but only a tiny effect on how serious his crime was thought to be. This is as expected: ancestrally (as now), antecedent circumstances provided ecologically valid cues to the future association value of an exploiter. In some sense, it is not surprising that people use past behavior to gauge future behavior. What is more surprising in the light of previous research is, first, the very weak effect on perceptions of crime seriousness and, second, that perceptions of future behavior and perceptions of crime seriousness had distinct effects on other judgments. Perceptions of the criminal's future value had a sizeable effect on subjects' preferences for reparative over punitive sanctions but a much smaller impact on judgments of how intense those sanctions should be. Perceptions of the crime's seriousness showed the opposite pattern: they had a sizeable effect on judgments of how intense sanctions should be but a much smaller effect on preferences for reparative over punitive sanctions.

5.2.3. Vignette 3: rape by in-group vs. out-group members

Past behavior is a straightforward and relatively proximate cue to the future behavior of a criminal. Is a criminal's association value also influenced by more distal cues? Whether someone is seen as a member of one's own social group or coalition should powerfully influence their association value: compared to out-group members, in-group members are more likely to be available as interactants, more willing to render help, and more restrained

about inflicting costs. Correspondingly, out-group members should be seen as caring less about the welfare of members of one's own group than in-group members do, leading to the inference that one is less likely to be helped and more likely to be exploited by out-group members. In-group favoritism, coupled to out-group indifference or hostility, is one of the best documented phenomena in psychology (Brewer 1979; Yamagishi, Jin, & Kiyonari 1999).

To see whether in-group status affects association value, thereby affecting reparative sentiments, a third vignette was created. It described a case of rape and contained a between-subjects experimental design that manipulated the in-group status of the offender. In one condition, the rapist was described as a young man of immigrant background (a potential out-group member). In the other condition, the rapist was described as a young man doing his compulsory military service (an in-group member). Because people may differ with respect to the extent to which they view immigrants as in-group cooperators or out-group exploiters, we assessed the average effects of immigrant status, as well as the interaction between immigrant status and the subject's hostility toward immigrants (see online supplemental materials for measure, available on the journal's website at www.ehbonline.org). The results are shown in Fig. 3 (data analysis was parallel to that for Vignette 2).

Does perceived association value predict preferences for reparative over punitive sanctions? Yes. The preference for reparative over punitive strategies was correlated with the perceived association value of the criminal (b=0.29) with an effect size similar to that found for the first two vignettes (0.27, 0.40).

Perceptions of the seriousness of the crime were not correlated with reparative sentiments for either the rape or the robbery vignettes; seriousness had a small but significant effect on reparative sentiments only for vandalism, a nonserious, nonviolent crime. This suggests that variations in the perceptions of the crime's seriousness predict reparative sentiments only for a restricted class of crimes: those that are seen as inflicting very little harm.

Does a distal but ecologically reasonable cue of association value—whether the criminal is an in-group or out-group member—contribute more strongly to perceptions of association value than to perceptions of the seriousness of his crime? Yes. The effects of the criminal's in-group status were tested using both additive and interactive regression models. In the additive model (reported under "All"), we tested whether the criminal's in-group vs. out-group status feeds into an index of the crime's seriousness or an index of the criminal's association value without considering whether the subject is hostile toward immigrants. In support of Prediction 3, out-group status significantly decreased the criminal's perceived association value but had no significant effect on perceptions of how serious the crime was.

Notice that the effect of group membership on perceived association value is weak compared to the effect of cues about past behavior found for Vignette 2. This is expected, in part, because association value was operationalized as the probability of future criminal behavior: in-group status is a much more distal cue of probability of future criminal behavior than the criminal's criminal history. In addition, however, the weak effect is expected because people differ with respect to the extent to which they see immigrants as in-group cooperators or

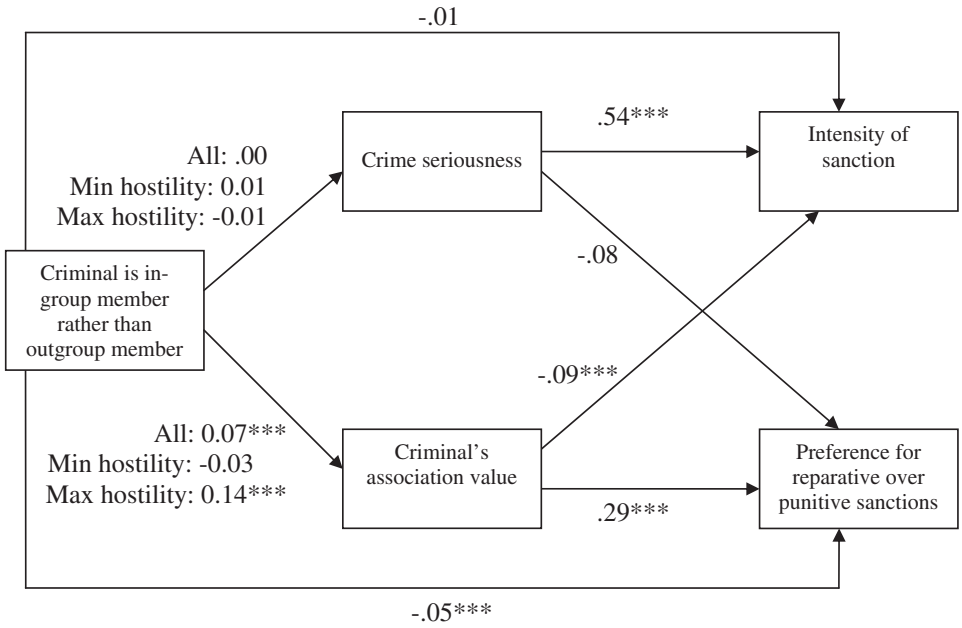

Fig. 3. Vignette 3, Study 1 (rape): perceived seriousness of crime and perceived association value of the criminal as mediators of a specific criminal's group membership. Unstandardized regression coefficients (b): all variables are coded between 0 and 1. Coefficients for individuals with minimum and maximum out-group hostility are calculated based on a model with a two-way interaction term between out-group hostility and experimental condition.

out-group exploiters, and such differences will weaken the effect of the out-group manipulation.

To provide a more realistic picture of the effects of manipulating in-group status—one that takes into account different attitudes toward immigrants—we analyzed the effects of a two-way interaction term between the experimental manipulation and the subjects' general out-group hostility (hostility toward immigrants in this particular case). As predicted, the interaction term had a significant effect on perceptions of the criminal's association value (b=0.17, p<.001) but no effect on perceptions of the seriousness of his crime (b=−0.02, p=.25).

Calculations of the marginal effects of the manipulations for the least and most hostile subjects are presented in Fig. 3 (under "Min out-group hostility" and "Max out-group hostility"). As these figures reveal, out-group status significantly decreased the criminal's association value among those most hostile toward immigrants (b=0.14, p<.001) but had no significant effect on the perceptions of the seriousness of his crime. Hence, Prediction 3 is again supported. Association values are indeed computed based on cues distinct from those used to compute a crime's seriousness.

Does in-group/out-group status affect preferences for reparative over punitive sanctions? If so, is this mediated by its effects on perceptions of the criminal's association value? We first consider in-group/out-group status independent of the subject's hostility toward immigrants. For the additive model, a Sobel test demonstrates that the mediation effect of association value on reparative sentiments is significant (z=6.30, p<.001). Again, to provide a more realistic picture of the effects of manipulating in-group/out-group status, we also analyzed the two-way interaction effect between the experimental manipulation and the subjects' hostility toward immigrants. As expected, this interaction term also influenced preferences for reparative over punitive sanctions (b=0.10, p=.002; i.e., as out-group hostility increased, subjects became more reparative towards the in-group member relative to the out-group member), but this direct effect became nonsignificant once association value was accounted for (b=0.05, p=.15). That is, the interactive effect of immigrant status and immigrant hostility on reparative sentiments was fully mediated by how these factors influenced perceptions of association value.

Does the perceived seriousness of the crime predict the preferred intensity of the sanction? Yes. Preferred intensity of sanction was predicted by both the perceived seriousness of the crime (b=0.54, p<.001) and perceptions of the criminal's association value (b=−0.09, p<.001). However, the effect size for seriousness was six times greater than for association value (it was four times and two times larger in Vignettes 1 and 2, respectively). Again, the difference between the absolute effect sizes is highly significant (F=207.10, p<.001).

The results of Vignette 3 support the hypothesis that association values and seriousness regulate different aspects of the counterexploitation system. Higher association values predicted a preference for more reparative strategies and less intense sanctions; moreover, their positive association with reparative strategies was six times greater than their negative association with sanction intensity. Conversely, higher seriousness values predicted preference for more intense sanctions but were uncorrelated with preferences for reparative strategies.

6. Study 2

To provide additional evidence for the recalibrational theory of counterexploitation and how it influences lay intuitions about crime, we analyzed data from a second study. This study allowed us to replicate the findings from Study 1 in another cultural context: the United States rather than Denmark. It used a measure of association value that taps the expectation that valuable associates are likely to provide benefits in the future, and a sanctioning preference measure that distinguishes more sharply between punitive and reparative strategies. Study 2 also varied cues to association value: half the subjects were told the criminal felt remorse, had no criminal record, and was a local businessman (a cue to in-group status); the other half were not provided with any information regarding these factors.

6.1. Methods

In Study 2, we tested our predictions by analyzing subjects' responses to a vignette with an embedded experiment that was designed to manipulate the association value of the criminal. The experiment was embedded in a Web survey given to undergraduates at a large American research university. The experiment involved the participation of 138 subjects. Given this smaller sample size, the directionality of our hypotheses and the supporting results from Study 1, one-tailed t tests were used to test all of the hypotheses in Study 2. The online supplemental materials (available on the journal's website at www.ehbonline.org) provide a range of extra analyses that demonstrate the robustness of these tests.

In the experiment, subjects were presented with a case of aggravated battery involving a knife. In one condition, the subjects were presented with the details of the crime. In another condition, cues that should increase the offender's association value were included by adding that the criminal "who manages an area auto parts store, expressed deep remorse and apologized for the pain he has caused the victim and his family." Furthermore, it was said that when learning of the criminal's arrest, "local residents were shocked (…) saying that he had no prior history of causing trouble."

In response to the vignette, we obtained four basic measures, as in Study 1: (1) the subjects' perceptions of the criminal's association value, (2) their perceptions of the crime's seriousness, (3) whether they preferred a reparative or punitive sanction, and (4) the preferred intensity of the sanction. Descriptions of all survey items appear in the online supplemental materials (available on the journal's

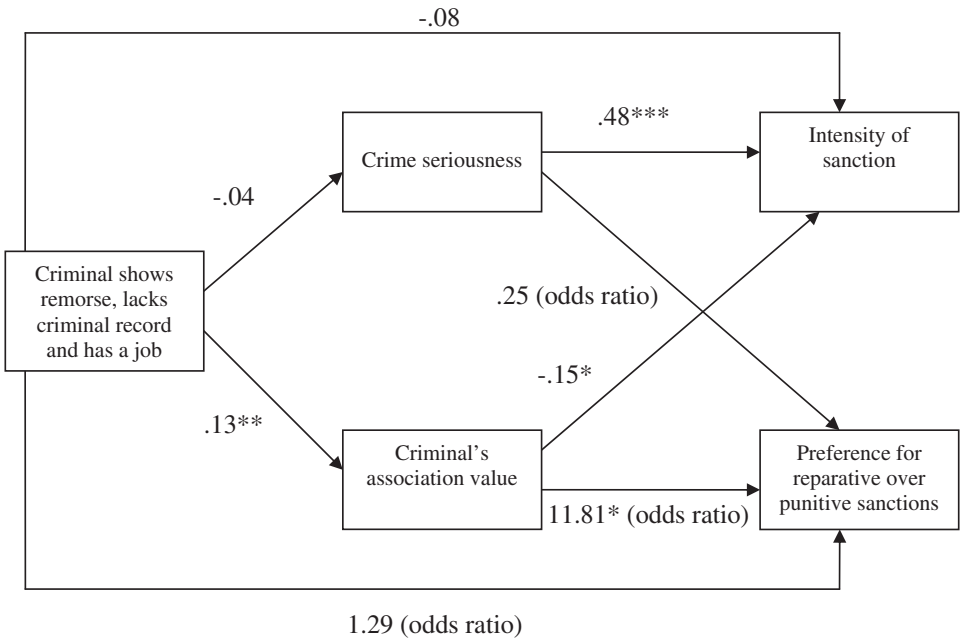

Fig. 4. Study 2 (battery): perceived seriousness of crime and perceived association value of the criminal as mediators of cues to association value. Unstandardized regression coefficients (*b*) and odds ratios calculated using binary logistic regression. All variables are coded between 0 and 1.

website at www.ehbonline.org). Here, we describe two changes in measurement compared to Study 1.

First, in Study 1, we measured association value using a measure that tapped the expectation that the criminal will impose costs in the future. In Study 2, we obtained a similar measure. However, a high association value should also lead to the expectation that the person might provide benefits in the future. To tap this expectation, the subjects in Study 2 were also asked how likely it is that "this criminal can someday become a productive member of society." The main text reports the effects of this new productivity measure, while the online supplemental materials (available on the journal's website at www.ehbonline.org) provide analyses of the recidivism measure for Study 2.[2] Second, in Study 1, we measured preference for a reparative vs. a punitive sanction by asking what the goal of the sanction should be. In Study 2, however, we asked whether the subject would recommend that the criminal participate in a rehabilitation program (job training/college degree and/or drug and alcohol treatment) instead of serving a prison sentence. Based on the answers to these questions, we created a dichotomous measure of punitive vs. reparative preference. Subjects who chose rehabilitation over prison at least once were coded as having a reparative preference (coded 1). Subjects who never chose rehabilitation over prison were coded as having a punitive preference (coded 0). This measure of the preference for a punitive (prison) vs. a reparative (rehabilitation program) sanction allowed us to provide a more stringent test of the recalibrational theory: Do subjects sanction criminals with higher and lower association not just with different goals in mind but also with markedly different kinds of sanctions?

In Study 2, we follow the same strategy for data analysis and presentation as in Study 1. The only difference is that in the case of reparative over punitive preferences, we are using a dichotomous dependent variable and therefore use binary logistic regression rather than OLS regression and, as is standard, report odds ratios. The results are shown in Fig. 4.

6.2. Results

Does perceived association value predict preferences for reparative over punitive sanctions? Yes. As shown in the bottom right of Fig. 4, the perceived association value influences the choice between a reparative over a punitive strategy (odds ratio=11.81, p=.03, one-tailed). The odds ratio of 11.81 implies that a change in the perceived association value of the criminal from the variable's minimum value (0) to its maximum value (1) makes it 11.81 times more likely that the subject will choose a reparative rather punitive sanction. This finding replicates the relationship between association value and reparative preferences found in Study 1 while measuring association value as expectations of becoming a productive member of society rather than expectations of committing crimes in the future. Furthermore, as shown in Fig. 4, crime seriousness did not have a significant effect (odds ratio=0.25, p=0.49).

[2] In the online supplemental materials, it is demonstrated that the recidivism and productivity measures are highly correlated (and likely track the same underlying psychological variable) and the effects reported here are replicated using a scale combining the two measures.

Do cues of remorse, lack of criminal record, and employment contribute more strongly to perceptions of association value than to perceptions of the seriousness of his crime? Yes. As shown in the left side of Fig. 4, variation in information about the criminal had a significant effect on perceptions of his association value (b=0.13, p=.005, one-tailed), but no effect on how serious his crime was seen to be (r=−0.04, p=.25, one-tailed). The effect of the cues on perceived association value was three times greater than their effect on the perceptions of the crime's seriousness.

To test whether these cues affect the preference for a reparative over a punitive sanction indirectly, through their effects on perceptions of the criminal's association value, we performed a formal mediation test suited for binary dependent variables (Stata FAQ 2012). We find that perceived association value does significantly and fully mediate the effect of the experimentally varied cues on sanctioning preference.[3]

Does the perceived seriousness of the crime predict the preferred intensity of the sanction? Yes. As shown in the top left of Fig. 4, the preferred intensity of sanction was predicted by both the perceived seriousness of the crime (r=0.48, p<.001, one-tailed) and perceptions of the criminal's association value (r=−0.15, p=.03, one-tailed), but the effect size for seriousness was more than three times larger than for association value.

These results provide strong further support for the recalibrational theory counterexploitation. We were able to replicate the core findings from Study 1 in a different country using a benefit-oriented measure of association value—one that does not tap expectations about future criminal activities—and a more demanding measure of reparative preferences.

7. Discussion

Extending extant research on lay intuitions about criminal justice, we have demonstrated that individuals spontaneously compute an index of the criminal's association value and that this value regulates the motivation to repair the criminal or punish him. Across a range of different types of crime and across two highly different countries (the United States and Denmark), subjects' preferences for rehabilitation over punishment were regulated by their perceptions of the criminal, independently of their perceptions of the crime. The seriousness of the crime, as judged by the subjects, did not regulate their preferences for repair over punishment for violent crimes. The effect of seriousness on reparative sentiments was significant only for the vignette describing a nonviolent crime, vandalism; even so, its effect size was only half that for the criminal's association value. The seriousness of the crime, in contrast, regulated the intensity of preferred sanctions far more than perceptions of the criminal did.

A person's association value is an estimate of how likely that person is to exploit (or confer benefits) on you and those you care about in the future. We experimentally manipulated several ancestrally valid cues to association value: the offender's criminal history, the offender's status as an in-group or out-group member, and the offender's expression of remorse. The results confirmed that these cues have a major effect on the computation of a criminal's association value but little or no effect on computations of the crime's seriousness. As previous studies have shown, the seriousness of the crime seems instead to be computed primarily on the basis of the costs the crime imposes on others (see online supplemental materials, available on the journal's website at www.ehbonline.org). These results support the hypothesis that the mind's design for deciding how to respond to criminals (exploiters) has two distinct information processing channels, which use two distinct sets of cues to compute two different decision variables: the criminal's association value and the seriousness of the crime. In sum, these results show that while a crime's indexed seriousness regulates how much to react, the criminal's indexed association value regulates the more fundamental decision of how to react (i.e., whether we want to punish or rehabilitate).

Today, there is little chance that an individual's personal welfare will be affected by whether the state punishes or rehabilitates a specific criminal or even fails to react at all. Nevertheless, our intuitions seem to reflect a strategic social calculus that operates as if crimes occurred in an intimate social setting where we ourselves required punitive protection or could harvest the social value of a repaired relationship. This calculus, we have argued, mirrors the adaptive problems posed by the detection of exploiters in the small-scale, interdependent social environments of our hunter/gatherer ancestors. In a world without modern technology or welfare systems, social support from resourceful others was a significant survival factor, and natural selection has favored designs that trigger reparative over punitive strategies in response to exploitive acts by associates who are, or might become, valuable to us. These selection pressures have left their imprint on the human mind, causing modern individuals to reason flexibly about the sanctioning process, as if it occurred in a small-scale setting.

Deontological notions of justice are not sufficient to account for these lay intuitions about criminal justice: our results show that preferences for rehabilitation over punishment are regulated by our perceptions of the criminal's future behavior—his or her value as a future associate—independent of our judgments about the seriousness of the crime that was committed. By implication, disagreement on the appropriate sanction against a criminal will often occur within a community. The same transgressor may have elicited a different association value in the mind of each member of the community: the transgressor might be of the same ethnicity as me, but an out-group member to

[3] The coefficient of the indirect path is 0.08. All values of the 90% confidence interval for this mediation effect (i.e., corresponding to the one-tailed test for the existence of a mediation effect) are positive (0.01–0.20), indicating a significant effect at the .05-level (one-tailed).

you; a kinsman of mine, but not of yours; a cooperative partner of mine, but a competitor of yours; and a longstanding ally of mine, but an enemy of yours. Given that each individual's intuitions about criminal justice are a function of the distinct association value that he or she computed for the transgressor in question, disagreement within and across communities on the appropriate sanction is the direct implication of—rather than evidence against— the role of evolved sentiments in criminal justice intuitions.

Supplementary Materials

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.evolhumbehav.2012.05.003.

Acknowledgments

This research has been presented at the Research Section for Comparative Politics, Aarhus University and at the Center for Evolutionary Psychology, UCSB. We thank participants for constructive inputs and feedback. Also, we are very grateful to Elizabeth Suhay for invaluable help in collecting data for Study 2.

References

Aharoni, E., & Fridlund, A. J. (2011). Punishment without reason: isolating retribution in lay punishment of criminal offenders. *Psychology, Public Policy, and Law*, Advance online publication.

Aureli, F., & de Waal F. (Eds.), (2000). *Natural conflict resolution.* Berkeley and Los Angeles: University of California Press.

Boehm, C. (1985). Execution within the clan as an extreme form of ostracism. *Social Science Information*, *24*, 309–321.

Braithwaite, J. (2002). *Restorative justice and responsive regulation*. New York: Oxford University Press.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: a cognitive–motivational analysis. *Psychological Bulletin*, *86*, 307–324.

Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature*, *373*, 209–216.

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, *83*, 284–299.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Cermak, I., & Rosza, S. (2001). Facing guilt: role of negative affectivity, need for reparation, and fear of punishment in leading to prosocial behaviour and aggression. *European Journal of Personality*, *15*, 219–237.

Cords, M., & Killen, M. (1998). Conflict resolution in human and non-human primates. In J. Langer, & M. Killen (Eds.), *Piaget, evolution, and development* (pp. 193–217). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cullen, F. T., & Gendreau, P. (2000). Assessing correctional rehabilitation: policy, practice, and prospects. *Criminal Justice*, *3*, 109–175.

Daly, M., & Wilson, M. (1988). Homicide. Hawthorne, NY: Aldine.

Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human Behavior*, *24*, 659–683.

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, *7*, 324–336.

de Waal, F. (1996). Good natured: the origins of right and wrong in humans and other animals. Cambridge, MA: Harvard University Press.

Delton, A. W. (2010). A psychological calculus for welfare tradeoffs. University of California, Santa Barbara: Unpublished dissertation.

Duntley, J. D. (2005). Adaptations to dangers from humans. In D. M. Buss, Ed. *The handbook of evolutionary psychology* (pp. 224–249). Hoboken, NJ: John Wiley & Sons.

Duntley, J. D., & Buss, D. M. (2004). The evolution of evil. In A. Miller, Ed. *The social psychology of good and evil* (pp. 102–123). New York: Guilford.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*, 63–87.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review*, *90*, 980–994.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Frank, R. H. (1988). *Passions within reason: the strategic role of the emotions*. New York: W.W. Norton & Company.

Fry, D. P. (2000). Conflict management in cross-cultural perspective. In F Aureli, & F. de Waal (Eds.), *Natural conflict resolution* (pp. 334–351). Berkeley and Los Angeles: University of California Press.

Fujisawa, K. K., Kutsukake, N., & Hasegawa, T. (2005). Reconciliation pattern after aggression among Japanese preschool children. *Aggressive Behavior*, *31*, 138–152.

Harris, N. (2003). Reassessing the dimensionality of the moral emotions. *British Journal of Psychology*, *94*, 457–473.

Harris, N., Walgrave, L., & Braithwaite, J. (2004). Emotional dynamics in restorative justice conferences. *Theoretical Criminology*, *8*, 191–210.

Jacoby, S. (1983). *Wild justice: the evolution of revenge*. New York: Harper & Row.

Kelly, R. L. (1995). *The foraging spectrum: diversity in hunter–gatherer lifeways*. Washington: Smithsonian Institution Press.

Ketelaar, T., & Au, W. T. (2003). The effects of feelings of guilt on the behavior of uncooperative individuals in repeated social bargaining games: an affect-as-information interpretation of the role of emotion in social interaction. *Cognition and Emotion*, *17*, 429–453.

Lieberman, D., & Linke, L. (2007). The effect of social category on third party punishment. *Evolutionary Psychology*, *5*, 289–305.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2011). Cognitive mechanisms for revenge and forgiveness. *The Behavioral and Brain Sciences*, in press.

Petersen, M. B. (2012). Social welfare as small-scale help: evolutionary psychology and the deservingness heuristic. *American Journal of Political Science*, *56*, 1–16.

Petersen, M. B., Sell, A., Tooby, J., & Cosmides, L. (2010). Evolutionary psychology and criminal justice: a recalibrational theory of punishment and reconciliation. In H. Høgh-Olesen, Ed. *Human morality and sociality: evolutionary and comparative perspectives* (pp. 72–131). Hampshire: Palgrave Macmillan.

Roberts, J. V. (2008). Punishing persistence: explaining the enduring appeal of the recidivist sentencing premium. *British Journal of Criminology*, *48*, 469–481.

Roberts, J. V., & Stalans, L. J. (2004). Restorative sentencing: exploring the views of the public. *Social Justice Research*, *17*, 315–334.

Robinson, P., Kurzban, R., & Jones, O. D. (2007). The origins of shared intuitions of justice. *Vanderbilt Law Review*, *60*, 1633–1688.

Rossi, P. H., Simpson, J. E., & Miller, J. L. (1985). Beyond crime seriousness: fitting the punishment to the crime. *Journal of Quantitative Criminology*, *1*, 59–90.

Selke, W. (1991). A Comparison of punishment systems in Denmark and the United States. *International Journal of Comparative and Applied Criminal Justice*, *15*, 227–242.

Sell, A. (2005). *Regulating welfare tradeoff ratios: three tests of an evolutionary–computational model of human anger*. PhD Dissertation. Santa Barbara: Center for Evolutionary Psychology, University of California, Santa Barbara.

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences*, *106*, 15073–15078.

Silk, J. B. (2002). The form and function of reconciliation in primates. *Annual Review of Anthropology*, *31*, 21–44.

Stata FAQ (2012). "How can I perform mediation with binary variables?." UCLA Academic Technology Services, http://www.ats.ucla.edu/stat/stata/faq/binary_mediation.htm, accessed April 22 2012.

Stylianou, S. (2003). Measuring crime seriousness perceptions: what have we learned and what else do we want to know. *Journal of Criminal Justice*, *31*, 37–56.

Tooby, J., & Cosmides, L. (1996). Friendship and the bankers paradox: other pathways to the evolution of adaptations for altruism. *Proceedings of the British Academy*, *88*, 119–143.

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions*, 3rd ed. (pp. 103–129). New York: Guilford Press.

Tooby, J., Cosmides, L., & Price, M. E. (2006). Cognitive adaptations for n-person exchange: the evolutionary roots of organizational behavior. *Managerial and Decision Economics*, *27*, 103–129.

Vangelisti, A. L., Daly, J. A., & Rudnick, J. R. (1991). Making people feel guilty in conversations: techniques and correlates. *Human Communication Research*, *18*, 3–39.

Yamagishi, T., Jin, N., & Kiyonari, T. (1999). Bounded generalized reciprocity: ingroup boasting and ingroup favoritism. *Advances in Group Processes*, *16*, 161–197.

Warr, M., Meier, R. F., & Erickson, M. L. (1983). Norms, theories of punishment, and publicly preferred penalties for crimes. *The Sociological Quarterly*, *24*, 75–91.