

Original Article

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

Michael Bang Petersen<sup>a,\*</sup>, Aaron Sell<sup>b,c</sup>, John Tooby<sup>b</sup>, Leda Cosmides<sup>b</sup>

<sup>a</sup>Department of Political Science and Government, Aarhus University, Aarhus, Denmark

<sup>b</sup>Center for Evolutionary Psychology, University of California, Santa Barbara, CA, USA

<sup>c</sup>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.

© 2012 Elsevier Inc. All rights reserved.

*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).

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

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات