BEHAVIORAL AND BRAIN SCIENCES (2013) **36**, 1–58
doi:10.1017/S0140525X11002160

Cognitive systems for revenge and forgiveness

Michael E. McCullough
Department of Psychology, University of Miami, Coral Gables, FL 33124-0751
mikem@miami.edu
http://www.psy.miami.edu/faculty/mmccullough

Robert Kurzban
Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, and Economic Science Institute, Chapman University, Orange, CA 92866
kurzban@psych.upenn.edu
http://www.psych.upenn.edu/~kurzban/

Benjamin A. Tabak
Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90095-1563
btabak@psych.ucla.edu

Abstract: Minimizing the costs that others impose upon oneself and upon those in whom one has a fitness stake, such as kin and allies, is a key adaptive problem for many organisms. Our ancestors regularly faced such adaptive problems (including homicide, bodily harm, theft, mate poaching, cuckoldry, reputational damage, sexual aggression, and the infliction of these costs on one's offspring, mates, coalition partners, or friends). One solution to this problem is to impose retaliatory costs on an aggressor so that the aggressor and other observers will lower their estimates of the net benefits to be gained from exploiting the retaliator in the future. We posit that humans have an evolved cognitive system that implements this strategy – deterrence – which we conceptualize as a revenge system. The revenge system produces a second adaptive problem: losing downstream gains from the individual on whom retaliatory costs have been imposed. We posit, consequently, a subsidiary computational system designed to restore particular relationships after cost-imposing interactions by inhibiting revenge and motivating behaviors that signal benevolence for the harmdoer. The operation of these systems depends on estimating the risk of future exploitation by the harmdoer and the expected future value of the relationship with the harmdoer. We review empirical evidence regarding the operation of these systems, discuss the causes of cultural and individual differences in their outputs, and sketch their computational architecture.

Keywords: adaptationism; aggression; computation; conflict; cost/benefit analysis; evolution; evolutionary psychology; forgiveness; function; punishment; reconciliation; social relationships; revenge; violence; social psychology

1. Introduction

1.1. Is revenge a "Disease"? Is forgiveness a "Cure"?

The desire for revenge is a cause of many forms of aggression (Carlson & Miller 1988; Richard et al. 2003). It is a causal factor in 10% to 20% of homicides worldwide (Carcach 1997; Daly & Wilson 1988; Dooley 2001; Gaylord & Galligher 1994; Kubrin & Weitzer 2003), 61% of school shootings (Vossekuil et al. 2002), and 27% of bombings (Bureau of Alcohol, Tobacco, and Firearms 1999). Moreover, the desire for revenge apparently makes people ripe for recruitment into terrorist organizations (Speckhard & Ahkmedova 2006).

Perhaps because the desire for revenge is so closely linked to violence, it has been fashionable in Western thought since the Stoic (and, later, Christian) philosophers to view revenge as immoral, irrational, or both (Jacoby 1983; Murphy 2003; Summerfield 2002). Social scientists in the past century also promulgated the idea that the desire for revenge is indicative of psychological dysfunction (Horney 1948; Murphy 2003; Summerfield 2002). Linking revenge to mental disorder seems reasonable at first glance because the desire for revenge is a common response to extreme violence and trauma, and because it is also associated with post-traumatic stress symptoms (Bayer et al. 2007; Orth et al. 2006; Parkes 1993).

Two decades ago, clinical psychologists and therapists endorsed (at least tacitly) the "revenge as disease" conceit as they initiated the psychological study of revenge's conceptual foil – forgiveness. The earliest published professional articles on forgiveness were descriptions of forgiveness-based therapeutic techniques for helping people recover from the effects of traumatic experiences and vengeful feelings on their psychological and relational functioning (e.g., Hope 1987; Human Development Study Group 1991; Marks 1988; Moss 1986; Phillips & Osborne 1989; Pingleton 1989; Ritzman 1987; Worthington & DiBlasio 1990). Likening revenge to a disease has had a predictable effect on how forgiveness has come tacitly to be understood: If the desire for revenge is a disease, then perhaps forgiveness is the cure.

For instance, many of the earliest empirical studies on forgiveness were related to the use of interventions for

 0140-525X/13 $40.00

promoting forgiveness in therapeutic settings (DiBlasio & Benda 1991; Freedman & Enright 1996; Hebl & Enright 1993; McCullough & Worthington 1995), and much of the scientific literature on forgiveness implies that forgiveness, as an alternative to revenge, has positive consequences for human health and well-being (Worthington et al. 2007). The clinical interventions that have emerged from scholarly interest in the links of forgiveness to health and well-being are generally effective at promoting forgiveness – as well as at reducing psychological symptoms of anxiety and depression and boosting self-esteem (Baskin & Enright 2004; Lundahl et al. 2008); and revenge is often linked negatively (and forgiveness positively) to indicators of physical and mental health (Worthington et al. 2007). Relatedly, some researchers characterize forgiveness as a salutary alternative, but one that is also difficult to enact and easily disrupted by constraints such as poor executive function (Pronk et al. 2010), temporary depletions of psychological resources that are necessary for self-control (Dewall et al. 2007), or symptoms of mental disorder (Orth et al. 2008).

Such claims do not in themselves, however, license the view that revenge is best likened to a disease and forgiveness to a difficult-to-implement cure. Suppressing coughing, sneezing, and other symptoms might make patients feel better; however, these might be best thought of as normal and functional aspects of the body's defenses (Nesse & Williams 1994).

MICHAEL E. MCCULLOUGH is a Professor in the Department of Psychology at the University of Miami, Florida, where he directs the Evolution and Human Behavior Laboratory. His research has focused on a variety of social phenomena in humans, including forgiveness, revenge, religious belief and behavior, gratitude, and self-control. He is a Fellow of Division 8 (Society for Personality and Social Psychology) of the American Psychological Association (APA), and in 2000 he won the Margaret Gorman Early Career Award from APA's Division 36 (Psychology of Religion). His most recent book is *Beyond Revenge: The Evolution of the Forgiveness Instinct*, published in 2008 by Jossey-Bass.

ROBERT KURZBAN is an Associate Professor in the Department of Psychology at the University of Pennsylvania. He has published numerous journal articles on an array of topics drawing on his background in evolutionary psychology and behavioral economics, including morality, cooperation, friendship, mate choice, supernatural beliefs, modularity, and self-control. He serves as co-Editor-in-Chief of *Evolution and Human Behavior*, and in 2008 he won the inaugural Distinguished Scientific Award for Early Career Contribution from the Human Behavior and Evolution Society. His first book, *Why Everyone (Else) is a Hypocrite: Evolution and the Modular Mind*, was published in 2010 by Princeton University Press.

BENJAMIN TABAK is an NIMH Postdoctoral Fellow in the Department of Psychology at the University of California, Los Angeles. He received his Ph.D. in Clinical Psychology from the University of Miami in 2011. He has published several journal articles and book chapters on biopsychosocial factors related to interpersonal conflict and its resolution, including: "Oxytocin indexes relational distress following interpersonal harms in women," published in 2011 in *Psychoneuroendocrinology*.

1.2. An alternative model: Evolved mechanisms for revenge and forgiveness

However, other theoretical approaches to understanding revenge and forgiveness are possible – and, indeed, are commonly used in the biological sciences. In this article, we propose that revenge and forgiveness result from psychological adaptations that became species-typical because of their ancestral efficacy in solving recurrent social problems that humans encountered during evolution (Williams 1966). Revenge and forgiveness, we argue, have complementary biological functions: We posit that mechanisms for revenge are designed to deter harms, and that forgiveness mechanisms are designed to solve problems related to the preservation of valuable relationships despite the prior impositions of harm.

Our goals here are to (a) define revenge and forgiveness in functional terms that will make them more amenable to an adaptationist analysis (Williams 1966); (b) describe the selection pressures that give rise to systems for revenge and forgiveness; (c) explain cultural and individual differences; and (d) outline the proximate causes and the computations involved when these systems are performing their evolved functions.

2. Defining revenge

2.1. Non-functional approaches to defining revenge

To appreciate the benefits that might come from conceptualizing revenge and forgiveness in functional terms, it is useful to start by considering some of the definitions that have guided previous scholarship on revenge. Govier (2002), for example, wrote, "When we seek revenge, we *seek satisfaction by attempting to harm the other (or associated persons) as a retaliatory measure*" (p. 2, emphasis in the original). Elster (1990) likewise defined revenge as "the attempt, at some cost or risk to oneself, to impose suffering upon those who have made one suffer, because they have made one suffer" (p. 862). Uniacke (2000) also claims that "revenge is personal and non-instrumental: with revenge we seek to make people suffer because they have made us suffer, not because their actions or values require us to bring them down" (p. 62). Social psychologists, too, often use "the intention to see the transgressor suffer" (Schumann & Ross 2010, p. 1193) as a key definitional element of revenge.

These and other definitions (e.g., Carlsmith et al. 2008; Frijda 1994; McCullough et al. 2001; Mocan 2008) all capture the notion that revenge is harm imposed in response to some triggering violation or infliction of harm, but these proximate explanations leave a promissory note for an ultimate explanation that must be paid (Tinbergen 1963). *Why* should revenge produce pleasure? "Enjoyment" is not a complete explanation for behavior, but is rather an important part of the phenomenon to be explained (West et al. 2011). Is revenge an adaptation and, if so, what fitness benefits explain its existence (Andrews et al. 2002; Tooby & Cosmides 1992; Williams 1966)?

2.2. Defining revenge functionally

A functional definition of revenge addresses these issues. Biologists sometimes define behavior functionally. For example, Maynard Smith and Harper (2003) defined a *signal* as "any act or structure which alters the behaviour of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved" (p. 3). Likewise, West et al. (2007) defined cooperation as "a behaviour which provides a benefit to another individual (recipient), and which is selected for because of its beneficial effect on the recipient" (p. 416). To an evolutionary biologist or evolutionary psychologist, the function of a behavioral mechanism is the effect that causes the mechanism that produces that behavior to evolve (Andrews et al. 2002). By hypothesizing a function, it becomes possible to search for the behavioral or physiological features that contribute to accomplishing the putative function (Williams 1966). If the psychological systems that produce revenge (or any other behavior) do not show features supporting a hypothesized function, the hypothesis that a given system serves the hypothesized function is undermined.

2.2.1. A function for revenge: Changing other individuals' incentives toward the self. We think revenge solves an adaptive problem that is faced by many species: how to change other organisms' incentives to emit benefits and to avoid imposing costs upon oneself (Clutton-Brock & Parker 1995; Daly & Wilson 1988; Tooby & Cosmides 1996). Specifically, we hypothesize that cognitive mechanisms for revenge evolved because their behavioral outputs (i.e., retaliatory impositions of costs or withholdings of benefits) caused individuals to revise downward the net returns they expect to receive by engaging in exploitive behaviors against the vengeful individual in the future, which in turn (a) deters them from efforts to exploit the retaliator (Bshary & Grutter 2005) or (b) induces them to emit benefits for the sake of the retaliator. Cognitive systems that motivate organisms to provide these sorts of contingent punishments and rewards (Seymour et al. 2007) may boost their bearers' lifetime reproductive fitness (Chagnon 1988; but cf. Beckerman et al. 2009) and thereby evolve for their ability to change other organisms' incentives toward the self (Tooby & Cosmides 1996). Our approach to revenge has an affinity with Sell et al.'s (2009) recalibrational theory of anger, which claims that anger is an evolved motivational program designed to motivate behavior in the angry individual that will cause the individual at whom the anger is directed to revise upward the value he or she places on the angry individual's welfare.

2.2.2. Welfare tradeoff ratios. To further develop our proposal that revenge is designed to raise individuals' incentives toward the self, we use Tooby et al.'s (2008) Welfare Tradeoff Ratio (WTR) concept. Tooby et al. (2008) conceptualize WTRs as internal regulatory variables, stored in memory and continually updated, that humans use to guide social decision making according to appropriate criteria. An individual's WTR for a target individual reflects how much the bearer values the target's welfare – operationally, how large a benefit to the self the bearer would be willing to take if doing so required imposing a given cost of upon the target. WTR values of 1:1 indicate that the actor values the welfare of the target individual equally to his or her own: The actor would impose a 1-unit cost upon the target individual if and only if it, in turn, led to at least a 1-unit benefit for the self. WTRs less than 1 – say, 1:2 – indicate that the actor values the target individual's welfare one-half as much as one's own, and therefore, that the actor would willingly impose a cost of up to 2 units upon the target to obtain one unit of benefit or more for himself or herself (Petersen et al. 2010; Sell et al. 2009; Tooby et al. 2008). A WTR of 0 indicates that an actor would willingly impose a cost of any size upon the target to obtain even an infinitesimally small benefit: such a target's welfare is completely ignored. Moreover, individuals will impose costs on enemies even if doing so is a cost, rather than a benefit, to themselves: Reflecting this reality of social behavior, WTRs can be negative (Petersen et al. 2010).

WTRs, according to Petersen et al. (2010) and Tooby et al. (2008), exist in two natural kinds: monitored and intrinsic. An actor's monitored WTR for a target individual expresses the regard an actor should have for the target's welfare in light of the social consequences that the actor's behavior might trigger if the target (or that individual's kin, friends, or allies) were to discover the actor's behavior and then attempt to harm or help the actor on the basis of that information. Monitored WTRs are responsive to factors that modify a target's ability to respond retributively (i.e., with rewarding or punishing), such as physical strength, coalition size, and access to resources (Petersen et al. 2010). Monitored WTRs regulate actors' behaviors by way of systems that estimate the probability of contingent reprisals and systems that estimate the probability of contingent rewards from the targets of those behaviors.

In contrast to monitored WTRs, intrinsic WTRs express the regard that an actor possesses for a target individual's welfare solely due to the indirect effects of the actor's behavior toward the target on the actor's own welfare (rather than because of the target's ability to monitor and impose rewards or sanctions). Intrinsic WTRs are cognitive representations of the interdependence of the welfare of an actor and a target irrespective of any retributive responses that the actor's behavior might elicit from the target of that behavior. The claim that people can have different monitored and intrinsic WTRs for the same target individual explains how it is possible for humans to treat a feared coworker with kid gloves in the workplace, but then derogate the coworker in private. To know whether individual A's behavior toward individual B is being guided by A's monitored WTR toward individual B, one must observe how A's regard for B (as revealed through behaviors that influence B's welfare) changes as a function of whether B is able to learn of A's welfare-relevant action. This claim can also explain why humans behave with regard for their children's, mates', and friends' welfare even in situations in which the target individual could not possibly observe the actor's behavior.

Surrounding oneself with individuals who have high monitored WTRs toward the self (if monitoring is feasible) and high intrinsic WTRs toward the self (under all circumstances) is beneficial, as is distancing oneself from individuals who do not. In modern environments, however, many individuals in relatively close proximity consistently fail to provide benefits that they plausibly *could* provide. For

example, people could spontaneously offer strangers the contents of their wallets, but typically they do not. Generally, however, such non-deliveries of benefits neither cause humans to feel entitled to better treatment, nor to attack those that do not provide such benefits. Instead, humans form expectations for how other individuals should treat them based on relationship categories such as friend, father, mother, ally, stranger, enemy, dominant, and subordinate. Expectations also take into account other actors' behavioral histories, including direct experience and reputational information, as well as their perceived ability to harm and help (Petersen et al. 2010).

Assuming, then, that humans have computational processes for generating baseline WTRs (of both the monitored and intrinsic varieties) for other individuals – and estimates of those other individuals' baseline WTRs toward them – we also posit the existence of cognitive routines for registering that an actor has treated the self with less regard (i.e., that an actor has committed an action that connotes a lower WTR toward the self) than one would have expected based on one's previous estimate of the actor's WTR toward the self. At issue here is not simply whether an individual imposed a cost upon oneself, but whether that cost imposition was permissible given the victim's understanding of the harmdoer's WTR for the self. Parents, for example, readily abide their children's imposing many costs upon them over the life course, and few of those costs are met with revenge because those costs are perceived as permissible given not only the parents' high WTRs for their children, but also the parents' acceptance of their children's relatively low WTRs toward their parents. (Children's behavior, we would argue, is adaptively organized by cognitive mechanisms that motivate children regularly, and largely with impunity, to impose all sorts of costs – both large and small – on their parents. Breastfeeding, begging for food or attention, and tantrums are examples. Parents tacitly accept such costs even if they are inconvenient and annoying.) Understanding a harmdoer's intentions is important because accidental harm does not provide information about the actor's WTR. Intentional harm, however, implies that the harm was *caused by* the harmdoer's low WTR for the victim (Petersen et al. 2010; Sell 2011).

2.2.3. Revenge and WTRs. The introduction of the WTR concept here enables us to put our proposal concisely: The revenge system, by motivating retaliatory harm, was selected because it caused other individuals to raise their WTRs for avengers (see also Petersen et al. 2010; Sell et al. 2009) so that those individuals would refrain from imposing costs upon the avengers in the future. Imposing a large harm on a victim to obtain a small benefit for the self indicates that the aggressor does not highly regard the victim's welfare relative to the aggressor's own welfare. Revenge is an effort to compel an aggressor to increase his or her regard for the victim's welfare – essentially, to teach the aggressor that imposing costs of the same size upon the victim in the future (should they be detected) will be met with retaliatory costs. Again, the central logic is the logic of deterrence: If an aggressor learns that the victim will impose large retaliatory costs, then the aggressor can be made less likely to perform such acts in the future.

By "regard for the victim's welfare," we do not mean that revenge causes increased *feelings* of care or concern. Instead, we are arguing that revenge produces vigilance about imposing future costs upon the avenger. Thus, we define revenge as a *targeted imposition of costs or withholding of benefits, in response to a cost-inflicting (or benefit withholding) event, that results from a cognitive system designed (i.e., selected) for deterring other organisms from imposing costs (or inducing other organisms to confer benefits) upon oneself or other individuals in whom one has fitness interests.*

2.2.4. Other definitional considerations. Our definition of revenge is similar to many evolutionary biologists' definition of punishment. Many biologists have been heavily influenced by Clutton-Brock and Parker (1995), who defined punishment as a costly (in the currency of fitness) imposition of costs (in the currency of fitness) on another individual that results in delayed benefits (in the currency of fitness) for the punisher (see also Jensen 2010). Although our definition of revenge broadly matches Clutton-Brock and Parker's definition of punishment, we depart slightly from their approach: Our claim is that the imposition of costs at a cost is only revenge when it is a response to a harm-imposition or a benefit-withholding *that was caused by a mechanism designed to deter cost-impositions or benefit-withholdings in the future*. Revenge, by our definition, therefore differs from other forms of punishment such as those administered by individuals acting as representatives of social institutions (e.g., judges, school principals), and in the context of precautions or shaping (e.g., the adult who scolds a child for swearing, or the behaviorist who wants her laboratory rats to learn to avoid the first left turn in a maze).

Our attempt here is to model interactions among individuals rather than among groups of individuals. We take no position on whether the psychology that governs the operation of revenge systems (and the reconciliation systems we will discuss later) also evolved to regulate behavior in intergroup contexts. The computations required to make adaptive decisions about revenge (and forgiveness) in the context of intergroup conflict (e.g., see Lickel et al. 2006) might be different enough from those required to make adaptive decisions in the context of conflicts between individuals within a single living group as to require a distinct cognitive architecture (Petersen et al. 2010).

Before continuing, we clarify four points. First, when one's kin, friends, allies, mates, or offspring are harmed, one suffers indirect harm (in the currency of fitness), and the revenge system should reflect this fact (Lieberman & Linke 2007). Similarly, one can impose retaliatory harm by imposing costs on a provoker's kin, friends, allies, mates, and offspring (Aureli et al. 1992; Engh et al. 2005; Gould 2000). Our analysis of revenge applies to direct impositions of cost, as well as indirect costs that accrue via their effects on one's kin, mates, and affiliates (Aureli et al. 1992; Bernhard et al. 2006; Hamilton 1964; Shinada et al. 2004). In the remainder of this article, we do not always specifically add these indirect considerations in the interest of brevity. Second, and related, we intend for costs to be in the currency of fitness, even though the relevant psychology is not tracking fitness costs per se, but rather, appropriate proxies (e.g., somatic damage, damage to one's property, reputation, social relationships, etc.).

Third, we include as revenge many behaviors that do not, as a matter of fact, manage to deter anyone (as when people behave aggressively toward a driver whom they perceive to have mistreated them on the road, and with whom they will never interact again). Because the modern world consists of many one-shot interactions, in (possible) contrast to ancestral human environments (Hagen & Hammerstein 2006), systems designed to deter do not always implement their proper functions (Sperber 1994) – that is, the effects for which they were selected. Evolved mechanisms are not expected to perform the jobs for which they were naturally selected with perfect fitness-optimizing performance in every possible environment (West et al. 2011).

Fourth, harming a provoker is only revenge, we propose, when the system that motivated the harmful behavior was designed by selection pressures for deterrence. So, for instance, avoiding a provoker to avert a second harm is not revenge (instead, the harm imposed by avoidance, if any, might be a by-product of a mechanism designed to reduce the probability of future harm to the self), but avoiding a provoker for the purpose of limiting his or her access to a valued resource (i.e., to withhold benefits) might constitute revenge (Barnes et al. 2009). In subsequent sections, we situate avoidance as a strategy *per se* among a larger suite of strategies that includes avenging the harm, forgiving it, and ignoring it (and perhaps others).

3. A functional model of revenge

3.1. Selection pressures for a revenge system

Clutton-Brock and Parker (1995) noted that retaliation (which they called "punishment") is prevalent among nonhuman animals (for some examples, see Aureli et al. 1992; Hoover & Robinson 2007; Jensen et al. 2007; Silk 1992), and they hypothesized that retaliation produces fitness gains for a punisher by reducing the probability that the recipients of punishment will repeat their injurious actions against the punisher in the future. Likewise, we hypothesize, following Daly and Wilson (1988), that humans similarly have mechanisms designed to produce revenge that evolved because of their effectiveness in addressing this adaptive problem (see also Fitness & Peterson 2008; Petersen et al. 2010; Tooby & Cosmides 2008).

We envision four types of deterrence: (1) direct deterrence of cost impositions (i.e., harming the provoker directly); (2) indirect deterrence of cost impositions (i.e., harming someone whose fitness affects the provoker's own fitness); (3) direct deterrence of the withholding of benefits; and (4) indirect deterrence of the withholding of benefits. Here we describe the first three of these, omitting a discussion of the fourth because of the thinness of the empirical literature.

3.1.1. Direct deterrence of cost impositions. By direct deterrence, we mean that revenge discourages an aggressor (who imposed a cost upon an avenger) from imposing other costs upon the avenger in the future. The logic here, identical to the logic underlying deterrence theories of punishment (Bentham 1962), is that if a potential aggressor can choose whether to take an action that imposes costs on a potential victim to acquire some benefit, then the potential victim is better off if he or she changes the potential aggressor's incentives so that the expected value of the cost-imposing action for the potential aggressor is lowered. Revenge transforms expected value in this fashion by causing an aggressor to learn that the retaliatory infliction of fitness costs in the future is likely to exceed (or, at least, substantially reduce) the potential benefits to be gained by harming the potential victim.

To work as a deterrent, however, avengers must act in such a way that the aggressors update their expectations of the avenger's future behavior (and, in so doing, revise their monitored WTR toward the avenger, with the consequence that the harmdoer becomes more vigilant in avoiding imposing costs upon the avenger in the future). If an aggressor imposes a cost of 1 on a victim to obtain a benefit of 2, and the victim proceeds to impose a retaliatory cost greater than 2, this retaliatory response only acts as a deterrent if the aggressors updates his or her representation of future options to reflect the possibility that the victim will again impose (similarly sufficiently large) offsetting costs upon the aggressor in the future. Effective updating, of course, requires an array of computational steps: The target of revenge must, among other things, (a) store an internal regulatory variable (Tooby et al. 2008) that represents his or her level of regard for the avenger's welfare (this is Tooby et al.'s [2008] monitored welfare tradeoff ratio), (b) represent the magnitude of the costs and benefits, (c) represent the causal relationship between the initial aggressive act and the retaliatory act, and crucially, (d) infer that retaliation at one time point is diagnostic of how the avenger is likely to behave in similar future situations. As Petersen et al. (2010) point out, targets of revenge must also be able to make generalizations about the ranges of behaviors toward the avenger that are likely to be met with vengeance in the future.

The logic of revenge gives rise to strategic complications. For example, though revenge at time 1 *might* predict revenge at time 2, nothing forces this to be true. An organism could be, for example, intermittently vengeful. This leads to well-known problems associated with inducing others to learn that one's vengeful dispositions are stable over time (Frank 1988; Hirshleifer 1987). This idea is of interest in the field of international relations because nations must signal their willingness to take revenge, even during the end game of an armed conflict, if they wish to deter conflict in the first place (Schelling 1960).

Experimental evidence in support of the proposition that revenge is well-suited to deterring the imposition of costs comes from studies of human behavior in economic games such as the sequential and iterated Prisoner's Dilemma (Axelrod 1980; 1984). This literature is too large to summarize exhaustively, but several stylized findings are worth considering. In the sequential Prisoner's Dilemma Game there is one round of play, but the second mover chooses only after seeing the first player's choice. In such games, the second player is much more likely to cooperate after a cooperative move than after a defecting move. More relevant to our present point, defection is almost always met with retaliatory defection (see Table 6 in Clark & Sefton 2001; cf. Hayashi et al. 1999).

In the iterated Prisoner's Dilemma, subjects play multiple rounds of the simultaneous move game with either the same partner or different ones. For the present purpose, key issues are whether people respond to defection with defection – moves plausibly interpretable as revenge (though clearly open to other interpretations,

such as loss prevention) – and whether such moves elicit subsequent cooperation from one's partner. Experiments using large numbers of trials in Prisoner's Dilemma Games suggest that people do respond to defection with defection (Bixenstine & Wilson 1963) though the details vary across studies (Rapoport & Chammah 1965). Reciprocal strategies such as tit-for-tat tend to elicit cooperation (e.g., Axelrod 1984; W. Wilson 1971), hinting at the effectiveness of meeting defection with defection for eliciting subsequent cooperation (Gardner & West 2004).

Moreover, in an analysis of data from 5 different laboratory studies of dyadic negotiation in which partners played 250 consecutive trials in which they could either punish, reward, or withhold reward (and punishment) from each other, Molm (1997, see especially pp. 126–34) found that the frequency with which actors used retaliatory punishment (i.e., after one's negotiation partner had punished the actor in a previous move) was positively associated with the frequency with which they rewarded each other. Further, the partner of each dyad who used contingent punishment more frequently in response to his or her partner's previous cost imposition was the partner who received more benefit from it. In contrast, the frequency with which dyads punished *non-contingently* (that is, independently of whether the punishment was a retaliatory response to punishment or the withholding of benefits) was associated with lower rates of rewarding: It is only when punishment is contingent on previous punishment (or the withholding of benefits, as we elaborate in section 2.2.3) that it promotes cooperation.

In some situations, one can benefit from revenge's efficacy as a deterrent simply by advertising one's ability to retaliate; it is not always necessary to do so. The difference in how people play the Dictator Game, as opposed to the Ultimatum Game (Güth et al. 1982), illustrates this point. In both games, some amount of money, say $10, is to be divided between two people. In the Dictator Game, one person unilaterally decides how to split the money. In the Ultimatum Game, one person – the Proposer – proposes a split, and the other person – the Responder – can either accept the proposal or reject it, in which case both players receive nothing. When the Responder rejects the Proposer's offer, then, the Responder is penalizing the Proposer by the amount that the Proposer set aside for himself or herself, but to impose this penalty, the Responder also pays a price: the amount that the Proposer had allocated to the Responder.

Unsurprisingly, typical proposals in the Ultimatum Game, in which punishment is possible, are larger (roughly 40% of the stake; Oosterbeek et al. 2004) than they are in the Dictator Game (roughly 28%; Engel, in press), in which punishment is not possible. Along similar lines, Andreoni et al. (2003) compared the results of a Dictator Game to those of three other games in which the Responders could either (a) punish, (b) reward, or (c) both punish and reward. In the punishment condition, the second player could pay one unit to impose a five-unit cost upon the first player. Under such conditions, players specify larger transfers to the second player, though transfers are still higher when the second player has access to (using the authors' metaphor) both a carrot (i.e., increasing the first player's payoff) and a stick (i.e., the capacity to inflict retaliatory costs).

The Trust Game (Berg et al. 1995) shows a similar deterrent effect for the ability to punish. In the Trust Game, Player 1 – also sometimes called the "investor" – starts with an endowment of money and is given the opportunity to transfer some of it to Player 2, also known as the "trustee." Transferred money is multiplied (often tripled), and the trustee can then return some, none, or all if it back to the investor. Money sent by the investor is commonly interpreted as trust, and money returned is commonly interpreted as trustworthy behavior on the part of the trustee. Hopfensitz and Reuben (2009) used a game like the trust game in which the investors had only a binary choice – trust or not trust – and the trustees had only 3 choices – low, medium, and high levels of trustworthiness. When punishment is added to this game (by allowing investors and trustees alike to punish their partners, subject to certain conditions), investors are more likely to make the trusting choice, and trustees on average return more money back to investors. Here, the possibility of punishment changes behavior in the desired way: it increases trusting and trustworthy behavior. Fehr and List (2004) reported similar results with students and CEOs.

At least one laboratory experiment also shows how the risk of retaliation deters aggressors from harming the prospective avenger. Diamond (1977) had undergraduate men write an essay that a confederate proceeded to harshly criticize. Participants came back to the laboratory 24 hours later and were given the opportunity to give ten (bogus) shocks of varying intensities to the person who wrote the insulting reviews. Half were led to believe that after they administered shocks, they would then switch roles and receive shocks themselves. People who believed that they could harm the insulting evaluators without the threat of retaliation gave stronger shocks to the evaluators. Thus, the fear of retaliation deterred aggression. More generally, the possibility of retaliation has been used to explain why defections within cooperative systems (e.g., the relationships between cleaners and clients in cleaner-client mutualisms), and thus, punishment as well, is rare in many cooperative animal systems (Cant 2011).

3.1.2. Indirect deterrence of cost impositions. Psychological mechanisms for revenge might be designed to deter would-be aggressors, including those *who have not yet exploited the avenger*. The logic behind deterring third parties is parallel to that of deterring second parties. If a third party must decide whether to impose a cost on an individual, the prospective victim can change the third party's incentives if he or she previously demonstrated that when a previous provocateur took a similar course of action, he or she inflicted costs upon that provocateur. To the extent that the third party believes that his or her prospective victim will be consistent in his/her propensity for revenge – an important limitation – the third party will alter his or her choices accordingly. In other words, by knowing that an individual is prone to avenging costs that others have imposed upon him or her, third parties will learn to treat the avenger with greater care, and with less willingness to extract costs from the avenger, in the future (dos Santos et al. 2011).

The computational demands here are non-trivial. For revenge to induce learning in third parties, third parties must be able to categorize actions adequately to determine whether the aggressive acts they are contemplating against a prospective victim are sufficiently similar to previous

aggressors' actions against the prospective victim to draw the prospective victim's retaliation. Third parties must also be able to track the causal structure of their prospective victims' vengeful acts. Finally, third parties must be able to compute, for any range of benefits to be extracted from prospective victims, the retaliatory costs those prospective victims are likely to subsequently impose. Despite the computationally intensive nature of third-party deterrence, it is plausible that revenge functions to deter both potential repeat offenders and would-be first-time offenders in a way that parallels the legal distinction between special deterrence (directed at recidivism of a criminal offender, in particular) and general deterrence (directed at other possible violators). Psychological revenge systems can do double duty in the same way that criminal justice systems do.

Reputation, then, might play an important role in third-party deterrence (dos Santos et al. 2011). To the extent that ancestral humans lived in small, close-knit groups (Boehm 2008) without police, courts, and prisons for protecting individual rights, a readiness to retaliate might have been an important component of people's reputations that would not only benefit them within a single living group, but, more importantly, as they transferred from one small living group to another (Marlowe et al. 2011). Researchers have documented the importance of defense of honor, and the revenge that it stimulates, as a major cause of violence among people from many societies, including Mediterranean herding societies (Black-Michaud 1975), tribal Montenegro (Boehm 1987), urban white males in the Southern United States (Cohen et al. 1996; Nisbett & Cohen 1996), and disadvantaged urban African-Americans (Anderson 1999). A longitudinal study of roughly 900 adolescent boys also revealed that boys who endorsed street-code beliefs (e.g., that violence is an appropriate response to insults and violations of honor) went on one year later to engage in more violence, including greater participation in gang fights and attacks in which their goal was to seriously injure or kill someone (Brezina et al. 2004).

Laboratory research supports the notion that the psychological mechanisms that cause revenge are sensitive to the presence of third parties, which is consistent with the idea that revenge is enacted partly out of reputational concerns. Victims retaliate more when an audience has witnessed the provocation – especially if the audience communicates to the victim that he or she looks weak because of the harm suffered, or if the victim knows that the audience is aware that he or she has suffered particularly unjust treatment (B. R. Brown 1968; Kim et al. 1998). Moreover, when two men have an argument on the street, the presence of a third person doubles the likelihood that the encounter will escalate from a verbal altercation to one that involves violence (Felson 1982). Not all studies, however, find that the presence of observers increases victims' likelihood of punishing (e.g., Bolton & Zwick 1995).

3.1.3. Direct deterrence of the withholding of benefits. The logic of revenge applies as much to changing others' incentives to *deliver benefits* to the self as to simply refraining from imposing costs on the self: From the standpoint of natural selection, there is no principled difference between the two, even if it turns out to be the case that cognitive systems that track the delivery and omissions of desired or anticipated benefits are distinct from systems that track the imposition of costs. Obtaining benefits and avoiding costs are functionally identical (although registering the omission of benefits requires certain computations that registering the imposition of harms might not; see sect. 5.1).

In a study described in section 3.1.1 above, Molm (1997, pp. 126–34) found that the frequency with which actors used punishment in response to a partner's previous withholding of rewards was positively associated with the rate at which partners rewarded each other over the series of trials. In other words, a willingness to punish one's interaction partner in response to his or her withholding of benefits increased the partner's delivery of benefits.

Public Goods Games also illustrate how revenge can deter the withholding of benefits. In these games, a few (e.g., 4–6) participants receive initial endowments of money that they each can divide between two pools. One pool is private; only the subject benefits from money kept in his or her own pool. The other pool is shared; money placed in this pool is multiplied by a number greater than 1 and the product is then divided evenly among all group members. Money maximizers keep everything in their private pools; aggregate group wealth is maximized when everyone contributes their entire earnings to the public pool. These games are social dilemmas (Kollock 1998; Ostrom 1990) because they create a tension between individual and group outcomes and provide an assay of cooperation (for reviews, see Camerer 2003; Ledyard 1995). Several Public Goods Game experiments are particularly important to review here because they add an important dose of realism: The avenger must pay a cost for the opportunity to punish (Clutton-Brock & Parker 1995; Jensen 2010).

Yamagishi (1986) had subjects play 12 rounds of a Public Goods Game in 4-person groups. He varied whether participants could punish other members of the group, and varied the cost to reduce another player's payoff by one unit. Players used the system for administering punishment when the opportunity to punish was made available to them. When it was available, players contributed greater amounts to the public good (see also, Carpenter & Matthews 2004; Ostrom et al. 1992; Yamagishi 1988). These results imply that punishment reduces the withholding of benefits in these games.

From this and subsequent work, we can draw some tentative conclusions. Punishment seems to be particularly effective in eliciting contributions when (a) the punishing technology makes the cost of punishment relatively low in comparison to the cost imposed on the individual being punished (Egas & Riedl 2008; Nikiforakis & Normann 2008); (b) there are many repeated interactions (Egas & Riedl 2008; Gürerk et al. 2006; Walker & Halloran 2004); (c) people can communicate their intentions regarding investment levels and the use of punishment (Ostrom et al. 1992); and (d) people make choices about their preferred group partners on the basis of those prospective group members' contributions on previous rounds of play (Page et al. 2005), or can migrate into or out of groups that have the capacity to punish (Gürerk et al. 2006).

When a participant in such experiments punishes a group member who has withheld contributions to the public good, such punishment behavior could be caused by revenge systems, but it could arguably result from other systems – for example, one designed to produce behaviors that induce the targeted individual to emit

benefits for other people in the future, or one designed for the enforcement of social norms (Clavien & Klein 2010).

Fehr and Gächter (2002) ran an experiment in which players changed groups after every round so that punishment could not be used strategically to induce group members who were uncooperative in round r to cooperate with the punisher in round $r + 1$. Even with this methodological alteration, Fehr and Gächter (2002) obtained similar results as in Public Goods Game experiments in which participants played with the same partners in each game, or with individuals with whom they might be randomly paired in successive rounds (Fehr & Gächter 2000). Participants punished uncooperative group members, and group members cooperated more when the option of punishing was available to the group (see also Anderson & Putterman 2006). Fehr and Gächter (2002) interpreted these results as evidence for altruism rather than for revenge because punishers could not use punishment to help themselves directly: All players were regrouped after each round of play. We would argue, however, that a revenge interpretation of Fehr and Gächter's (2002) results is plausible (Clavien & Klein 2010; Kurzban & DeScioli, submitted). First, Fehr and Gächter (2002) describe subsidiary results based on participants' self-reported responses to hypothetical scenarios (see also Fehr & Fischbacher 2004) to make a case that third parties' anger (and other group members' fear of those third parties' anger) shapes cooperation and punishment decisions in Public Goods and Third-Party Punishment games. However, these data are consistent with a revenge interpretation also because anger is a common response to personal exploitation (Fessler 2010; Sell 2011; Sell et al. 2009; Srivastava et al. 2009). That is, we believe that anger is the motivational system that brings about revenge, though here it is not discharging its proper function (Sperber 1994): Our view is that punishment in the Public Goods Game is caused by a proximate psychology designed for deterrence of personal harms in a world in which interactions were generally repeated with a relatively small number of interactants (e.g., Hagen & Hammerstein 2006).

Moreover, other research traditions have associated empathy for victims, rather than anger toward perpetrators, as the proximate emotional cause of action whose goal is to deliver benefits to others (Batson 2011). Indeed, some work from this latter tradition suggests that individuals do not naturally become angry upon observing the mistreatment of one stranger by another stranger unless empathy for the victim was experimentally manipulated beforehand (Batson et al. 2007). Further, Shinada et al. (2004) found that (a) self-reported anger toward low contributors, and (b) judgments of the unfairness of low contributors' behavior are correlated with the extent to which one punishes low contributors within one's own groups, but they are not correlated with the extent to which one punishes low contributors in other people's groups. According to these latter experiments, anger evidently makes people punish individuals who have harmed them directly or who have disrupted cooperation in their groups, but in some circumstances, might not naturally motivate third-party punishment in response to witnessing someone receiving unfair treatment from a third party.

Fehr and Gächter (2002) construe their results as altruism (as opposed to revenge) because of the increase in contributions that punishment elicits from previously low contributors. They articulate the sense in which they mean that punishment is altruistic, writing that punishment "may well benefit the future group members of a punished subject, if that subject responds to the punishment by raising investments in the following periods. *In this sense*, punishment is altruistic" (p. 137, emphasis added). So, if punishment is being used in this sense, then punishment ought to decline as participants approach the end of the experiment, after which no one will engage in any more rounds of investment, and in one-shot games. This is contradicted by multiple studies (Anderson & Putterman 2006, see Footnote 8; Carpenter & Matthews 2004; Fehr & Gächter 2002; Page et al. 2005).

Our view doesn't commit to revenge instrumentally benefiting the vengeful individual in all instances. We take harm (or withholding of benefits) to be the eliciting factor, anger to be the proximate motivating system (Sell et al. 2009), and imposition of costs to be the behavioral output. So, although our view does commit to what will elicit revenge, the associated emotion of anger, and the behavioral output, we are not committed to the view that revenge will be absent when there is no chance of repeat interaction or that it will implement its proper function (e.g., Hagen & Hammerstein 2006; West et al. 2011). This argument could, of course, be applied equally if one were to argue that third-party punishment psychology (i.e., altruistic punishment) were also "misfiring" because it was designed for a world of repeat interactions (Anderson & Putterman 2006, see Footnote 8; Carpenter & Matthews 2004; Fehr & Gächter 2002; Page et al. 2005). Because of the way that payoffs are structured in Public Goods Games with punishment, revenge can generate benefits for others *as a side effect* (as in Fehr & Gächter 2002). The fact that some people benefit as a side effect of revenge should not necessarily cause one to infer that revenge systems are designed for altruism (Burnham & Johnson 2005; Kurzban & DeScioli, submitted; Price et al. 2002; West et al. 2011).

Some studies do show that some third parties punish individuals who fail to provide benefits to others even when those third-party punishers themselves have not been harmed, which suggests caution regarding a revenge interpretation. For example, Carpenter and Matthews (2012) ran a one-shot Public Goods Game and varied whether participants could punish members of their own groups (which could potentially influence the punishers' outcomes) or members of other people's groups (where the punishers' own welfare would be unaffected). In the key treatment, the "one-way TPP" (third-party punishment) condition (in which participants could punish individuals within their own groups and in other people's groups, but could not be punished by the individuals in other groups), they found that 90% of subjects did not punish outside their group (p. 12). The average amount participants used to punish third parties was roughly ten cents, and these same participants used approximately seven times as much money to punish low contributors within their own groups. Moreover, in a condition in which participants could only punish contributors within their own groups, the total expenditure for punishment was ten times as high as the expenditure for third-party punishment in the one-way TPP treatment.

Also, Carpenter and Matthews (2009) found in a repeated game that second-party punishment and third-party punishment did not differ greatly in magnitude. Further, some evidence indicates that people do engage

in third-party punishment in games other than Public Goods Games (Fehr & Fischbacher 2004), though the expenditures for third-party punishment are, as in Carpenter and Matthews (2012), lower than for second-party punishment. Nevertheless, in such games, the amounts of punishment that third parties mete out toward uncooperative or ungenerous individuals is not strong enough to deter individuals from withholding benefits from others, whereas second-party punishment is (Fehr & Fischbacher 2004). In any case, third-party punishment systems might or might not be designed to incentivize the targets of such punishment to benefit others in the future (DeScioli & Kurzban 2009b); for example, advertising one's willingness to engage in third-party punishment appears to produce reputational advantages (Barclay 2006; Nelissen 2008), which might suggest an alternative function (Yamagishi et al. 2009).

3.2. Cross-cultural commonalities and variability in the revenge system's operation

3.2.1. Cross cultural universality in the operation of the revenge system. Cross-cultural research suggests that revenge is a universal response (Brown 1991) to the imposition of costs – especially in the extreme case, the homicide of a kinsman. Ericksen and Horton (1992) found that 90% of the 186 societies in the Standard Cross-Cultural Sample (SCCS) showed clear evidence that blood feuds or individual self-redress were either actively used or formally outlawed in favor of formal adjudication procedures (the latter implying that revenge must have historically been a problem) "if a consanguineal kin group member is killed, injured, or insulted by a member of another kin group" (p. 60). For 18 of the 186 societies in the SCCS – most of the remaining 10% – data for making such a determination were either missing or conflicting. Relatedly, Daly and Wilson (1988) concluded that 57 (95%) of the 60 societies in the Human Relations Area Files probability sample had "some reference to blood feud or capital punishment as an institutionalized practice, or specific accounts of particular cases, or at the least, the articulate expression of the *desire* for blood revenge" (p. 226, italics in original).

More recently, Boehm (2008) reported results from an ethnographic survey of data for 10 "Pleistocene-appropriate" societies (i.e., economically independent pure hunter-gatherers that do not live in permanent year-round settlements). Boehm concluded that punishments of various types in response to prior harms were widespread, including public and private gossip about a violator (100% each); and physical punishment (90%).

A final line of cross-cultural evidence supporting the universality of revenge comes from Henrich et al.'s (2006) study of the Ultimatum Game in 15 small-scale societies (including two groups from North America, three groups from South America, six groups from Africa, a group from Asia, and three groups from Oceania) that differed in language, climate, and economic base. In all 15 societies, as proposers' offers tilted away from a 50/50 split in the proposers' favor, the recipients in all 15 societies became more likely to reject those offers. As Henrich et al. (2006) wrote, "In every population, less-equal offers were punished more frequently" (p. 1770). Taken together, the results from Ericksen and Horton (1992), Daly and Wilson (1988), Boehm (2008), and Henrich et al. (2006) suggest that revenge is widespread cross-culturally, if not indeed universal (Brown 1991).

3.2.2. Cross-cultural variation in the operation of the revenge system. Nevertheless, there is substantial cross-cultural and temporal variation in the overall amounts and forms of revenge. Across societies, the percentages of homicides attributable to revenge (for instance) range from as low as 8% to as high as 45% (e.g., Cardona et al. 2005; Daly & Wilson 1988; Gaylord & Galligher 1994; Kubrin & Weitzer 2003). If revenge is for deterrence, it might be more prevalent in cultural ecologies in which there are few or no institutions (police, courts) that deter interpersonal aggression and other forms of harm. (We note that we take the question of why and how these institutions emerge to be a separate issue.)

Ericksen and Horton (1992) coded 186 societies in the Standard Cross-Cultural Sample to determine which of three approaches was favored for settling grievances when a member of one's own kin group is killed, injured, or insulted by a member of another group: (a) kin group feuding (i.e., classic blood feuding), (b) individual self-redress (individuals avenging their grievances on their own), and (c) formal adjudication. They found that individual revenge was the primary means of redress in traditional foraging societies. When people organize into tribes – especially tribal societies that emphasize manly honor (e.g., Boehm 1987) – kin groups take up relatives' grievances. Second, kin feuding evidently gives way to formal adjudication as external political forces implement the rule of law. Third, revenge by individuals or extended kin groups is replaced by formal adjudication in societies with favorable resource bases and vertical inheritance – that is, societies in which individuals would stand to benefit from the social stability that comes from replacing revenge with other means of sanctioning. In brief, then, these findings suggest that when institutions arise to administer third-party punishment, individual acts of revenge are crowded out. (One could equally conceptualize this as institutions raising the price of revenge, thereby reducing the demand for it.)

Anderson (1999) has suggested that concentrated neighborhood disadvantage (e.g., low socioeconomic status [SES] and median family income) precipitates a widespread lack of trust in formal legal methods for settling differences, and instead encourages the adoption of a "code of the street" that prescribes the use of personal revenge, rather than appeals to law enforcement authorities for settling one's interpersonal grievances (Brezina et al. 2004). Research on contemporary geographic distributions of revenge homicides supports this view. Kubrin and Weitzer (2003) found that concentrated neighborhood disadvantage was an excellent predictor of retaliatory homicide (i.e., homicides motivated by a desire for revenge in response to a previous perceived slight or injury): Indeed, the census tracts in St. Louis, MO, with high levels of poverty, high unemployment rates, and high percentages of children not living with both parents also had the highest rates of retaliatory homicide. Conversely, in census tracts with very low concentrated neighborhood disadvantage, retaliatory homicide was negligible. (These data leave open the possibility that violence, generally, is more common in such environments, as opposed to revenge, narrowly, being more common.) This insight is echoed in many

other ethnographic studies that find that revenge is more frequently used in societies in which social institutions for settling grievances are generally viewed as weak, and in which individuals are socialized to defend their honor with retaliatory violence at even the most trivial of interpersonal slights (Black-Michaud 1975; Boehm 1987; Nisbett & Cohen 1996).

3.3. Individual differences in the operation of the revenge system

There are individual differences in (a) the stated desire for revenge after being harmed (Singer et al. 2006); (b) the strength of the pattern of neural activation that correlates highly with the stated desire for revenge (Singer et al. 2006); and (c) the strength of retaliatory behavioral responses (Eisenberger et al. 2004). For example, women have moderately lower self-reported tendencies to seek revenge after being harmed than do men (Miller et al. 2008), and are less aggressive in response to provocation than are men (for an extensive review, see Bettencourt & Miller 1996). Given the marginal effectiveness of physical violence on the part of women compared to men, this is not surprising (Archer 2009; Sell et al. 2009).

Research in behavior genetics suggests that roughly 30–40% of the variance in individual differences in people's (self-reported) propensities to seek revenge results from additive genetic effects (Eaves et al. 2008). Shared environmental influences account for relatively little (i.e., approximately 15%) of the variance (Eaves et al. 2008). The remaining variance (45–70%) is attributable to non-shared environment, non-additive genetic effects, gene–environment interactions, and measurement error. Our claim that there is a species-typical, evolved mechanism for revenge is not, of course, undermined by evidence of substantial individual differences (Buss 2009; Tooby & Cosmides 1990a). Indeed, if revenge is a system for deterrence, then the system should be sensitive to characteristics of the would-be avenger that would render revenge more or less effective as a deterrent for that individual. Revenge (in particular, imposing physical harm on others) does not return equal benefits, or exact equal production costs, across all individuals (Sell et al. 2009). For those individuals for whom the costs are too high and the benefits too low, alternatives to revenge should be preferred. More generally, it is likely the case that the operation of species-typical psychological mechanisms is influenced by the other characteristics of the individual in which those mechanisms reside (or, as considered in sect. 3.2.2, the environments of the individuals in which those mechanisms reside). Applying this reasoning to what is currently known about the correlates of individual differences and their genetic and environmental substrates helps clarify the literature on individual differences in revenge.

3.3.1. Genetic sources of individual differences. Efforts to isolate genetic markers associated with individual differences in revenge have begun (McDermott et al. 2009), although the source of variation (e.g., random, unselected noise, frequency dependent selection for individual differences in vengefulness; environmental heterogeneity in fitness optima; see Buss 2009) remains unknown. One possibility is that these genetic individual differences can be thought of as contingent shifts in social strategy as a function of other heritable phenotypic characteristics – a phenomenon that Tooby and Cosmides (1990a) called *reactive heritability*.

Suppose the costs of revenge are higher, on average, for people with low upper body strength or body size – traits that are highly heritable (Carmichael & McGue 1995; Silventoinen et al. 2008) – because the vengeful efforts of weak individuals are more likely to be answered with counter-revenge (Sell et al. 2009). The same logic explains why men with greater body weight and height (both of which are indicative of physical formidability) are more likely to be diagnosed with antisocial personality disorder (Ishikawa et al. 2001), why large football players are perceived as more aggressive, less friendly, and less cooperative than are smaller players (Koenig & Ketelaar 2006), and why men's physical strength is positively associated with their aggressiveness, anger-proneness, their histories of fighting and success in conflict, and their beliefs about the utility of personal aggression (Gallup et al. 2007; Sell et al. 2009). If this supposition is correct, then statistically equating men on physical strength should substantially reduce the contribution of additive genetic factors to individual differences in the propensity for revenge.

The same logic builds a causal bridge between the fact that 99.9% of women have less upper body strength than does the average man (Lassek & Gaulin 2009) and the facts that (a) women are nearly a standard deviation less vengeful than men are (Archer 2009; Miller et al. 2008); (b) women are half a standard deviation less physically aggressive than men are, despite being no less anger-prone than men are (Archer 2004); and (c) sex differences in provoked aggression are strongest in experimental situations in which females are at greater risk than males of becoming the targets of counter-aggression (Bettencourt & Miller 1996). For women, the costs of revenge may best be lowered not through taking revenge via their own physical strength, but through either (a) indirect aggression such as reputational damage (Hess & Hagen 2006), or, as Sell et al. (2009) suggest, (b) recruiting coalitional support from others who can effect revenge on their behalf and (with their physical strength) deter counter-retaliation.

Phenotypic factors associated with women's success in recruiting male coalitional support might include factors such as physical attractiveness or waist-hip ratio that relate to mate value – some of which are highly heritable in women (Olson et al. 2001; Zillikens et al. 2008). Indeed, Sell et al.'s (2009) findings that women's physical attractiveness, but not physical strength, predicts their anger-proneness, beliefs about the utility of personal aggression, sense of entitlement, and self-reported history of success in resolving interpersonal conflicts in their own favor suggests that women's ability to leverage male coalitional support lowers the costs of revenge for women. If this explanation is correct, then heritable individual differences in women's vengefulness should shrink after statistically controlling for measures of mate value.

3.3.2. Environmental sources of individual differences. As with other social strategies, the propensity for revenge might vary across individuals because of the costs and benefits of using the strategy (Buss 2009). Cultural influences such as concentrated neighborhood disadvantage (see section 3.2.2 above), civic trust, policing efficiency,

and even parental support for retaliation as a way of handling grievances (Black-Michaud 1975; Brezina et al. 2004; Copeland-Linder et al. 2007; Ericksen & Horton 1992; Herrmann et al. 2008; Kubrin & Weitzer 2003; Nisbett & Cohen 1996; Solomon et al. 2008) are good candidates for explaining variance in revenge. Variables such as endorsement of a street code of conduct (Brezina et al. 2004; Stewart et al. 2006), and (mis)trust in the police (Kääriäinen 2007) may also be useful for capturing important aspects of the proximate psychology through which such environmental effects influence the propensity for revenge.

Finally, there is substantial individual variance in revenge that is attributable to non-shared environmental factors – that is, factors that monozygotic twins do not share in common. To the extent that revenge produces deterrence or other beneficial effects such as an improved reputation, increased social status, or increased attractiveness to prospective mates (Anderson 1999; Boehm 1987), it necessarily produces these effects with respect to *specific* bullies, despots, friends, peers, and potential mates, and these unique effects will calibrate individuals' propensities to engage in revenge in unique ways. Non-shared environmental effects have been notoriously difficult to identify in behavioral-genetic research (Turkheimer & Waldron 2000) because they are virtually infinite in number and generally small in magnitude. Nevertheless, as these unique social experiences accumulate over the life course, they will make even identical twins increasingly different from each other (Harris 2006). The cumulative effect of these unique social experiences will be to alter people's computations of revenge's costs and benefits, thereby yielding different propensities for revenge over the life course.

4. The evolution of forgiveness

4.1. Revenge-based costs and design for forgiveness

Revenge carries costs that can potentially offset its deterrence benefits. Although the costs of the act of revenge can sometimes be small – for example, spreading gossip or injuring a much smaller individual – these costs can sometimes be large – for instance, (to use a contemporary example) suicide bombing. Even in two-person Prisoner's Dilemmas (Dreber et al. 2008; Rand et al. 2009; Wu et al. 2009) and Public Goods Games (Bochet et al. 2006), the costs of punishment are often large enough to negate any gains in payoffs that punishment produces by increasing partners' cooperation.

In most cases, however, a more important cost of taking revenge lies in the fact that other people also have revenge systems, so costs imposed on them might cause them (or their kin, friends, or allies) to engage in counter-revenge (Boehm 1987; Cinyabuguma et al. 2006; Denant-Boemont et al. 2007; Dunbar et al. 1995; Gould 2000; Herrmann et al. 2008). Research on the iterated and sequential Prisoner's Dilemma illustrates the general point. A move of defect when both players are using a "tit for tat" strategy (Rapoport & Chammah 1965) locks players in a cycle of mutual defection known as the "echo effect" (Axelrod 1984), which drastically reduces payoffs. Because of this effect, "forgiving" strategies (e.g., responding to a partner's defection on round t by cooperating on round $t+1$) such as "generous tit for tat," "contrite tit for tat," and "firm but fair" (Frean 1994; Hauert & Schuster 1998; Nowak & Sigmund 1993; Wu & Axelrod 1995) reduce the chance of getting trapped in defect-defect spirals. Indeed, when noise is present, such strategies elicit more cooperation from human cooperators than does tit for tat (Bendor et al. 1991; Klapwijk & Van Lange 2009; Van Lange et al. 2002). Vendettas and blood feuding illustrate this point ethnographically (Boehm 1987).

A third cost of revenge applies when the aggressor is someone of value to the victim, such as a friend, genetic relative, or close ally. In such cases, taking revenge carries the cost of damaging one's own interests indirectly. For this reason, we expect revenge to be less frequently imposed upon kin, people with whom one has an ongoing exchange relationship (Trivers 1971), friends and allies (DeScioli & Kurzban 2009b), and long-term mates (Clutton-Brock 1989). Related, but importantly different – and arguably the principal cost relevant to revenge systems – is that revenge runs the risk of turning a friend into a foe, in which case the expected downstream value of the relationship is sacrificed, or flipped from positive to negative. Although new relationships can replace old ones, establishing new positive social relationships – including search and the accumulation of trust – represents a cost (Hruschka & Henrich 2006).

4.2. Reducing the costs of revenge

The costs of revenge, then, depend on the nature of relationship between the aggressor and the victim. To explore this important point, we again refer to the "welfare tradeoff ratio" concept (Petersen et al. 2010; Sell et al. 2009). Recall from section 2.2 earlier that imposing a large harm upon a victim to obtain a small benefit for the self indicates that the aggressor does not highly regard the victim's welfare relative to his or her own. Revenge, therefore, is a way for victims to attempt to cause their aggressors to increase their regard for the welfare of their victims.

Deciding whether to take revenge, then, should reflect a computation that weighs the expected benefits of revenge (e.g., will it cause the aggressor to update, in the favorable direction, his or her monitored WTR toward the victim?) against its costs (e.g., will the aggressor or his or her allies engage in counter-revenge, or update his or her monitored WTR toward the victim in the unfavorable direction?). The benefit side of the computation might be similar for both friends and foes: The key consideration is whether the act of revenge will deter future cost impositions upon the victim. However, when the aggressor is a friend or ally, the estimates of the costs must incorporate additional terms. As we have argued, there are the additional indirect costs associated with imposing harm on one's relatives and allies and the potential harm to existing mutually profitable relationships. We posit, therefore, the existence of mechanisms whose function is to inhibit revenge when the costs of revenge outweigh its deterrent benefits and to steer organisms toward other approaches to up-regulating aggressors' WTRs toward the self. These are *forgiveness systems*.

One factor in this computation is the aggressor's relationship to the victim. People should be less likely to take revenge on friends and relatives because of their fitness interdependence (Dunbar et al. 1995), which is expressed computationally as the victim's intrinsic WTR toward the aggressor. We are not claiming that people forgive all

harms done to them by genetic relatives: Conflicts could arise between relatives due to, for example, parent–offspring conflict (Trivers 1974). We nevertheless expect that there are systems designed to take kinship or other social factors that influence fitness interdependence into account in this context. Second, to the extent that revenge will reduce or eliminate the benefits to be captured from one's kin, friends, allies, mates, and so on, a well-designed system should assess the value of the relationship (see sect. 5.3 in this article, and Petersen et al. 2010), and raise the other's WTR toward the self without, if possible, imposing costs. We believe that forgiveness systems have this function: *Raising an aggressor's WTR toward the self without using revenge* (the active imposition of retaliatory costs) *to do so*. After Petersen et al. (2010), we believe that the function of forgiveness systems is to up-regulate the aggressor's WTR toward the victim by motivating the victim to behave in ways that will raise the aggressor's *intrinsic WTR* (but not the monitored WTR) toward the victim. Recall from section 2.2.2 that intrinsic WTRs reflect one's willingness to take benefits for oneself at a cost to target individuals based on all of the factors that create fitness interdependence between the actor and the target. When an individual forgives, he or she is attempting, therefore, to establish positive relations with a harmdoer by first causing that harmdoer to increase his or her intrinsic valuation of the victim.

4.3. Forgiveness: A functional definition

When a victim simply refrains from retaliation – for example, when a dominant individual has exploited a much weaker subordinate – forgiveness is not implicated in the sense in which we intend (cf. Gardner & West 2004). When revenge is not taken in such instances, it might be because, for example, one is physically unable to do so, or because revenge will invite even more exploitation in the future.

When friends, kin, or allies (rather than strangers, rivals, or enemies) have harmed the self, however, the relational costs associated with revenge (as described in sect. 4.2) also apply. In such instances, the revenge system can be in conflict with the putative forgiveness system. In previous work, the first author's research group has defined forgiveness as a set of motivational changes whereby an individual becomes (a) less motivated to retaliate against an aggressor; (b) less motivated to maintain estrangement from an aggressor; and (c) more motivated by good will for the aggressor (McCullough & Root 2005; McCullough et al. 1997; 1998; 2003).

Here we suggest that forgiveness systems are designed to guide victimized individuals toward behaviors that will change aggressors' intrinsic WTRs toward the self without the use of retaliatory impositions of costs – specifically, by inhibiting revenge and by signaling one's view of the harm one has incurred, as well as one's willingness to return to constructive relations conditional on the aggressor refraining from similar cost impositions in the future (i.e., contingent on an updating of the aggressor's WTR toward the victim). This construal of forgiveness permits the conceptual distinctions that other theorists (e.g., Enright & Coyle 1998; Worthington 2005) consider important (e.g., that forgiveness is different from forgetting an offense, condoning it, or attempting to minimize its significance).

Choosing an appropriate behavior following a harmful act, then, requires that one consider, at a minimum, (a) the aggressor's WTR for the victim as implied by the harmful act; (b) the potential downstream benefits embodied in the relationship if it were to continue; and (c) the potential effectiveness of several behaviors that might be deployed to modify the aggressor's WTR toward the self. Different combinations of values for the variables implied in those considerations can produce many distinct behavioral options. In section 4.4, we limit ourselves to discussing four of them. After sketching these four options, we continue by discussing reconciliation, by which we mean the restoration of relations between aggressor and victim at mutually acceptable WTRs between offender and victim. Finally, we outline what we believe to be some critical computational steps that a well-designed cognitive architecture for generating adaptive behavioral responses to impositions of interpersonal harm (viz., revenge and forgiveness) must execute.

4.4. Forgiveness among a suite of other behavioral options

4.4.1. Acceptance. First, after an individual has imposed a cost upon the self, one might simply *tolerate, or refrain from responding to, the harm* (Gardner & West 2004) – essentially accepting the aggressor's WTRs for the victim as implied by the cost-imposing action. In such cases, victims simply absorb the costs they have incurred at the hands of the aggressor, avoid the costs associated with enacting revenge, and continue interacting with the aggressor without attempting to modify the aggressor's WTR toward the self. In such cases, social interaction between the aggressor and the victim continues, possibly with the victim lowering his or her estimate of the aggressor's WTR toward him or her. Such a course of action might occur when the costs of the harm are lower than the expected costs associated with attempting to up-regulate the aggressor's WTR toward the self. Specifically, we predict that individuals tend to ignore harms whose costs to the self (discounted by the benefits to be obtained from deterring future similar harms by the aggressor or third parties who might be deterred indirectly) are lower than the costs associated with attempting to adjust the aggressor's WTR toward the self. Acceptance may be signaled using language or by appeasement gestures that communicate one's willingness to accept certain costs imposed by the aggressor and an absence of any residual motivation to engage in retaliatory aggression.

4.4.2. Revenge. Second, one might take revenge – that is, attempt to impose a cost upon the aggressor with the goal of altering the aggressor's monitored WTR toward the self and to obtain the benefits of deterrence more generally. As a result of revenge, social relations might end (e.g., social relations might be completely terminated, or revenge might incapacitate or kill the aggressor), or counter-revenge might ensue. Alternatively, relations between aggressor and victim might be restored under renegotiated and mutually tolerable WTRs. We expect revenge when the benefits of deterring future harms and

adjusting the aggressor's WTR toward the self outweigh the costs associated with imposing a retaliatory response.

4.4.3. Avoidance. Third, one might reduce or terminate one's interactions with the aggressor – that is, render it more difficult for the aggressor to impose costs upon, or obtain benefits from, the victim. Avoidance reduces the likelihood that the aggressor will be in a position to impose costs upon the victim again in the future. Avoidance might be more likely to be chosen when (1) the likely effectiveness of revenge is low (for any of the reasons we have already described) and (2) the estimate of the residual value in the relationship is low. When avoidance evolves for its efficacy in deterring harmdoers by depriving them of benefits they could have acquired through cooperative interaction with the individual who conditionally avoids harmdoers – rather than solely for its self-protective effects – then avoidance is better classified as an exit-based form of revenge (Barnes et al. 2009; Cant & Johnstone 2006).

4.4.4. Forgiveness. Fourth, the victim might forgive – that is, attempt to raise the aggressor's WTR (particularly the intrinsic WTR) toward the victim without imposing costs or withholding benefits from the aggressor in a retaliatory fashion. Putative forgiveness systems coordinate several tasks: They inhibit (i.e., down-regulate the activity of) systems that motivate revenge and avoidance, and they motivate "reparative behaviors" (Petersen et al. 2010) which signal that (a) the aggressor's behavior damaged the victim, and that (b) despite that harm, there is the possibility of future gains from interaction if the aggressor is willing to refrain from similar aggressive actions in the future (i.e., increases his or her WTR toward the victim). To cause aggressors to recalibrate their WTRs toward their victims, forgivers might (among other things) attempt to remind aggressors of previous benefits that they have provided to the aggressors (Petersen et al. 2010). Forgiveness should be more likely as the value of the relationship to the transgressor goes up and as the value of any deterrence to be obtained through acts of revenge or avoidance goes down (e.g., when a transgressor has indicated a disinclination to impose similar of costs on the victim in the future; see sect. 5.3 further on).

4.5. Reconciliation as a relational outcome

Subsequent to acceptance, revenge, avoidance, or forgiveness, aggressors and victims might *reconcile*, which we take, following Worthington (2005), to be a process by which an aggressor and a victim communicate to one another that they have arrived at mutually acceptable WTRs toward each other that will govern future interactions. In some cases, reconciliation might follow confirmation that one partner's previous WTR for the other was too low (i.e., when the aggressor acknowledges that he or she inappropriately exploited the victim), or too high (i.e., when a victim acknowledges that the aggressor has – and will continue to have – a lower WTR for the victim than the victim previously believed), or just right (i.e., the victim accepts the aggressor's prior act was, in fact, within the bounds of the aggressor's WTR toward the victim: In such an instance we might say that the victim overestimated the implications of a cost-imposing behavior for his or her estimate of the aggressor's WTR toward the self). Reconciliation is, then, the termination of individuals' efforts to recalibrate one another's WTRs and the return to social relations at mutually endorsed WTRs (after Petersen et al. 2010).

There is a direct analog between our conceptualization of reconciliation and the concept to which the term *reconciliation* refers in the animal behavior literature. Many group-living animals (including many primates, some canids, and at least one corvid) engage in *conciliatory behavior* – friendly post-conflict interactions with conspecifics (Aureli & de Waal 2000; de Waal & van Roosmalen 1979). Conciliatory behavior following an aggressive interaction tends to be followed by reductions of aggression (Aureli & Schaffner 2002; Aureli & van Schaik 1991b; Castles & Whiten 1998b), increases in friendly contact (Koyama 2001), and reductions in post-conflict anxiety (Aureli 1997; Aureli & van Schaik 1991b; Castles & Whiten 1998b; Koski et al. 2007), all of which suggest that these conciliatory behaviors – for instance, grooming (Cheney & Seyfarth 2007) – function to reduce retaliatory aggression and foster a return to cooperative interaction. In this sense, these behaviors appear to facilitate reconciliation inasmuch as nonaggressive, cooperative interaction often follows from them.

5. Exploitation risk and relationship value: Two computations regulating revenge and forgiveness

Calculating the costs and benefits associated with the various courses of action one might take after another individual has imposed an unacceptable cost upon oneself requires some intermediate computations (see also Petersen et al. 2010) – namely, estimation of the risk that the aggressor will harm the victim again in the future (which increases the likelihood of revenge and reduces the likelihood of forgiveness), and estimation of the future value of the relationship with the aggressor (which reduces the likelihood of revenge and increases the likelihood of forgiveness). In this section, we elaborate on the computation of future exploitation risk and future relationship value, and describe some of the social factors that may be used as information by the cognitive mechanisms that execute these intermediate computations.

5.1. Computing risk of future exploitation

Well-designed systems for adaptive choice between revenge and forgiveness must consider the provoker's ability and intention to impose costs upon the victim in the future (see also Bentham 1962). In the limiting case, suppose that after an offense that occurred completely privately (to rule out the possibility third parties could learn of the offense), a highly valuable relationship partner could persuasively signal that he or she would never – or could never – again inflict such costs. In such a case, revenge would yield no deterrent benefit. Estimates that future similar harms are unlikely should, to some extent, inhibit revenge.

Consistent with this hypothesis, people more readily forgive transgressors whose behavior was unintentional, unavoidable, or committed without awareness of its potential negative consequences (Eaton & Struthers 2006; Fehr et al. 2010; McCullough et al. 2010), presumably because

such information reveals that the harmdoer's harmful behavior toward the victim does not reflect a propensity to harm the victim again in the future (Petersen et al. 2010). Information relevant to the transgressor's future intent can come from explicit acknowledgments of wrongdoing (Eaton et al. 2006), efforts to repay or undo the costs imposed upon the victim (Bottom et al. 2002; Zechmeister et al. 2004), or both (Eaton & Struthers 2006). Verbal expressions of sympathy for a victim's suffering and explicit declarations of one's intention to refrain from harming the victim in the future are influential aspects of effective apologies (Gold & Weiner 2000; McCullough et al. 1997; Nadler & Liviatan 2006; Zechmeister et al. 2004).

Effective apologies make transgressors seem more remorseful (Risen & Gilovich 2007), less blameworthy (Zechmeister et al. 2004), and higher in personality traits such as agreeableness, sincerity, compassion, kindness, genuineness, and dependability (Risen & Gilovich 2007; Struthers et al. 2008; Tabak et al. 2012), all of which appear to convey a lack of motivation to inflict costs upon the victim in the future. Indeed, admissions of guilt without corresponding efforts to compensate the victim or communicate remorse can actually inhibit forgiveness (Allan et al. 2006; Zechmeister et al. 2004). From the viewpoint of the model we are advancing here, such findings make sense because admitting culpability should strengthen victims' confidence in their beliefs that the harmdoer intentionally harmed the victim and might be disposed to behave similarly in the future, and thus, that forgiveness (i.e., the inhibition of revenge and efforts to up-regulate an aggressor's WTR peacefully) would be imprudent.

Verbal apologies can be easily faked, however (Frank 1988) – hence, the "cheapness" of "cheap talk" – which should cause one to wonder why they matter at all. We anticipate, therefore, that systems for forgiveness will be sensitive to cues of sincerity – that is, cues that the individual offering the apology is not being deceptive. Credibility-enhancing displays (Henrich 2009), therefore, tend to be associated with effective apologies. For example, conciliatory behaviors such as delivering large repeated benefits (which King-Casas et al. [2008] winsomely named "coaxing"), or exposing oneself to harm by the victim (e.g., submission gestures; see Matsumura & Hayden 2006) require stronger internal commitments to improved relations – the putting of money where one's mouth is (Bottom et al. 2002), and as a result, appear to facilitate forgiveness better than apologies without behavioral signs of an internal commitment to improved future relations.

Likewise, facial displays such as blushing facilitate forgiveness after some transgressions (de Jong et al. 2003). The reliability of such displays may be on account of their relative unfakeability (Frank 1988). Indeed, Dijk et al. (2009) discovered that people judged those who blushed after transgressions as more trustworthy, sympathetic, and socially skilled than individuals who did not blush – some of the same inferences that people make about apologetic transgressors (Risen & Gilovich 2007; Struthers et al. 2008). Also, participants infer that people who blush are ashamed or embarrassed about their transgressions (Keltner & Buswell 1997) – which makes sense if blushes are, in fact, signals that harmdoers are aware of the fact that they behaved in a way that was inconsistent with their former (or current) welfare tradeoff ratio for the person or persons whom they harmed.

Also, revenge is unnecessary (holding aside its value as a third-party deterrent) when additional transgressions are impossible. When the aggressor's future capacity for violence has been removed (which legal theorists call incapacitation), for instance, avenging a harm may yield little additional deterrent value. In some ethnographic accounts, reconciliation rituals involve the surrender of weapons (e.g., Boehm 1987), which symbolize an unwillingness to engage in future cost-imposing behaviors. Such rituals, along with verbal communications, and credibility-enhancing displays (Henrich 2009) can be combined to make a highly persuasive signal of one's unwillingness to harm one's victim again in the future.

5.2. Computing the expected future value of the relationship with the transgressor

The expected future value of a relationship is computed, we hypothesize, in much the same way that it would be in contexts other than the aftermath of a transgression (Tooby et al. 2008). Because of the well-known principles of kin selection, close relatives are likely to be a source of benefits, and thus, we expect that kinship will facilitate forgiveness (Lieberman et al. 2007). Similarly, people with whom one has a close history of association (Tooby & Cosmides 1992; Trivers 1971), shared interests (Tooby & Cosmides 1996), similar values (Davis et al. 2009), and many opportunities for mutually beneficial transactions are good candidates for forgiveness because of the possibility of continued gains from association.

Research supports this hypothesis. McCullough et al. (2010) found that scores on a self-report measure of perceived relationship value (e.g., "I thought about the things I still like about our relationship.") predicted the rates at which people forgave during the 100 days after another individual had harmed them. The association between relationship value and forgiveness persisted even after controlling for participants' sex, feelings of closeness and commitment to their offenders, their ratings of the painfulness of the transgression, the transgressor's responsibility and intentionality in committing the transgression, and the extent to which the transgressor apologized and made amends. Also, priming people with the names of other individuals with whom they are close leads to increased self-reported inclination to forgive a variety of hypothetical offenses, increased accessibility of the concept of forgiveness, and reduced deliberation about whether forgiveness is an appropriate course of action (Karremans & Aarts 2007). Such findings complement those from previous studies showing that people are more inclined to forgive individuals to whom they feel close and committed (Finkel et al. 2002; McCullough et al. 1998) or securely attached (Kachadourian et al. 2004). We hypothesize that forgiveness is associated with variables such as closeness, commitment, and attachment because they index perceived relationship value. Note, however, that in relationships in which the shadow of the future is particularly long – that is, in which there are multiple opportunities for future interaction (after controlling for other, and perhaps better, measures of relationship value) – the judicious use of revenge could lead to long-term payoffs for the avenger inasmuch as those acts of revenge might deter many future cost impositions and might induce many future conferrals of benefits.

Here too, there is an analog in the literature on reconciliation among nonhuman animals: Conciliatory behavior is most common in "valuable relationships" – that is, relationships whose restoration would be expected to yield fitness payoffs to interactants (de Waal 2000; Silk 2002; van Schaik & Aureli 2000). In nonhuman animals, "relationship value" can come in a variety of currencies that influence reconciliation, including genetic relatedness (Call et al. 1999; Fraser & Bugnyar 2011; Katsukake & Castles 2001), mate value (Watts 1995), coalitional support (Cordoni & Palagi 2008), and grooming/aiding effort (Cordoni & Palagi 2008; Fraser & Bugnyar 2011; Koski et al. 2007; Preuschoft et al. 2002; Romero et al. 2009). The effects of particular forms of relationship value on conciliatory behavior have also been demonstrated recently in studies of chimpanzees (Fraser et al. 2010; Koski et al. 2007; Watts 2006), brown capuchin monkeys (Daniel et al. 2009), domestic dogs (Cools et al. 2008), Hamadryas baboons (Romero et al. 2009), sifakas (Palagi et al. 2008), wolves (Cordoni & Palagi 2008), and Assamanese macaques (Cooper et al. 2005).

6. Individual differences in forgiveness

As with revenge, there are individual differences in the extent to which people forgive harms they have incurred (McCullough & Hoyt 2002). Much (i.e., approximately 57%) of the variance in people's (self-reported) propensities to forgive results from additive genetic effects. The remaining variation is attributable to measurement error, unique environmental effects, non-additive genetic effects, gene–environment interactions, and other sources of influence that cannot be partitioned into additive genetic effects and shared environmental effects (Steger et al. 2007).

Research links individual differences in forgiveness with personality variables such as high agreeableness, low neuroticism, and religiosity (Mullet et al. 2005). As with revenge, people may possess heritable phenotypes that modify the costs and benefits of forgiving. Sex is one such trait, and women appear to score higher on self-reports of general tendencies to forgive than men do (Miller et al. 2008). Such a finding might be clarified in the future by considering particular relationship contexts in which forgiving might have higher benefits (or lower costs) for women than for men.

Likewise, we anticipate that the marginal benefits of forgiveness will be greater for people who lack social partners (Hruschka & Henrich 2006). People who are motivated to maintain scarce relationships and form new ones should be more willing – all else being equal – to forgive harms in the service of these efforts. Conversely, we might expect forgiveness to be less common in ecologies in which social relationships are short-lived, though we know of no data on this issue.

7. Summary

The desire for revenge and the ability to forgive are universal human psychological endowments (Boehm 2008; Brown 1991; Daly & Wilson 1988; McCullough 2008). In this article we have posited that revenge and forgiveness result from cognitive mechanisms that were designed to deter and to reduce the costs of deterrence while preserving valuable relationships, respectively. We have sketched the computational structure required for these putative functions, and discussed evidence that bears on these provisions as well as data surrounding relevant individual differences.

For some crucial questions about the revenge and forgiveness systems we have posited here, data are scant, and we could only speculate. To this point, the attention placed on revenge and forgiveness in the psychological literature has been somewhat limited and undertheorized. For instance, although we think that revenge plays a large and important role in human social relationships, our study of 9 recent textbooks in social psychology reveals a surprising dearth of coverage: The median number of pages on which "revenge" or "retaliation" are indexed is 0, with a mean of 0.78. Oddly – given that it has been taken seriously within the social sciences only recently – forgiveness has fared better in social psychology textbooks, with a median number of indexed pages of 1 and a mean of 1.22. Even so, researchers' abilities to shed light on forgiveness too have been limited, we think, by a failure to consider the functions that forgiveness might have evolved to serve within the human behavioral repertoire. We hope the adaptationist framework in which we have tried to situate the concepts of revenge and forgiveness will assist other researchers in formulating new direction for research in these areas.

ACKNOWLEDGMENTS

The authors wish to thank Jeffrey A. Carpenter, Daniel P. Hruschka, Phil McCabe, Karthik Panchanathan, Eric J. Pedersen, Michael Bang Petersen, James K. Rilling, Aaron Sell, and David White for their helpful comments on an earlier draft of this article. Work on this article was generously supported by grants from the John Templeton Foundation, the Center for the Study of Law and Religion at Emory University, the Fetzer Institute, and Grant #R01MH071258 from the National Institute of Mental Health to the first author.

Open Peer Commentary

Why so complex? Emotional mediation of revenge, forgiveness, and reconciliation

doi:10.1017/S0140525X12000490

Filippo Aureli[a,b] and Colleen M. Schaffner[a,c]

[a]*Instituto de Neuroetologia, Universidad Veracruzana, CP 91190 Xalapa, Veracruz, Mexico;* [b]*Research Centre in Evolutionary Anthropology and Palaeoecology, Liverpool John Moores University, Liverpool L3 3AF, United Kingdom;* [c]*Department of Psychology, University of Chester, Chester CH1 4BJ, United Kingdom.*

faureli@uv.mx f.aureli@ljmu.ac.uk
cm.schaffner@gmail.com
http://www.ljmu.ac.uk/NSP/98732.htm

Abstract: Humans have the cognitive abilities to implement the revenge and forgiveness systems hypothesized by McCullough et al., but the evidence suggests that simpler processes may underlie most revenge cases in humans and other animals. The mediating role of emotions can be at the basis of the flexibility needed in the hypothesized systems and the associated assessment of social relationships.

In this target article McCullough et al. present a fascinating view on revenge, forgiveness, and reconciliation. Whereas the existence of forgiveness in nonhuman animals may be controversial (and difficult to demonstrate), there is compelling evidence for revenge and reconciliation. In addition, the key function of revenge in imposing a retaliatory cost on aggressors has already been proposed for macaques (Aureli et al. 1992; cf. Clutton-Brock & Parker 1995). Thus, it seems that humans share both the form and function of revenge with other animals. Why then do McCullough et al. hypothesize that humans have evolved a cognitive system that implements the deterrence strategy, which they label a revenge system? Are McCullough et al. implying that animals, which show similar revenge patterns to those of humans, have the same cognitive revenge system? If so, humans did not evolve it but have adapted it from what a common ancestor already had in place. However, McCullough et al. are more likely implying that humans have evolved a unique cognitive system for revenge that is too complex for other animal species.

The revenge system presented in the target article is indeed complex. Whereas humans certainly have the cognitive abilities to implement such a system, is it really necessary to explain the majority of cases of revenge perpetrated by human beings? Couldn't a simpler mechanism based on emotional mediation be at the core of the patterns of revenge that have been reported for humans and other animals? There has been growing attention to the mediating role of emotions in humans (Frijda 1986; Panksepp 1989; Rolls 1995) and other animals (Aureli & Schaffner 2002; Aureli & Whiten 2003; Crook 1989; Lott 1991; Owren & Rendall 1997; Pryce 1996). An important function of emotions is motivating organisms to act (LeDoux 1996; Rolls 1990). In this respect, emotions interface between sensory inputs and motor outputs in a way that allows the individual to take a particular motivational stance (Aureli & Whiten 2003; cf. Tooby & Cosmides' [1990b] system of coordination), which then constrains its behavior for an appropriate amount of time (e.g., a longish period in the case of fear caused by sighting a snake; cf. Damasio 1994; Johnson-Laird & Oatley 1992).

The concept of relationship value is central to the forgiveness system hypothesized in the target article and to reconciliation between former opponents, as acknowledged by the authors. However, the expected value of social relationships does not need complex computation about the future. Most of our decisions are not taken based on improbable knowledge of the future, but are guided by probabilistic estimates based on past experience (Tooby & Cosmides 1990b). This is especially true for social intercourse as current behavior is affected, entirely or in part, by the individual's memory of past interactions (Aureli et al., in press; Hinde 1979; Seyfarth & Cheney 2012). The assessment of social relationships requires bookkeeping of the various interactions with the partner, computation of their relative frequencies, and conversion of their quality and associated information into a common currency. Such an assessment seems to be complex. However, emotions can play a critical role.

Emotional mediation has already been suggested to be at the basis of the assessment of social relationships (Aureli & Schaffner 2002). The emotional experience of an individual is certainly affected by the frequency and quality of previous interactions with group members (see Aureli & Schino [2004] for a review). Furthermore, emotional states may express a crucial integration of the information contained in the various interactions between two partners and may change over time depending on the interactions exchanged. The emotional experience can then be functionally equivalent to the aforementioned processes of bookkeeping, computation, and conversion needed for relationship assessment (Aureli & Schaffner 2002). The resulting emotional experience is partner-dependent. Thus, emotional differences can be at the core of the observed variation in social interactions reflecting the variation in relationship quality across partners.

Biological systems do not emerge ex novo as elegant solutions, but develop from pre-existing structures and therefore are constrained by their evolutionary past. Humans have the abilities for complex computations as required by the proposed revenge and forgiveness systems, but they usually rely on evolutionarily older systems. Quick and accurate decision-making is based on the exploitation of how information is structured in the social environment mediated by emotions (Gigerenzer et al. 1999). Similarly, emotions can serve as somatic markers that allow rapidly rejecting or endorsing certain options based on the reactivation of emotional states associated with previous experiences and permitting the individual to efficiently make a decision (Damasio 1994; 1996).

When revenge spreads from the two opponents to family members, as in mafia vendettas, it seems more cognitively demanding because the individuals involved need to know about the relationships of others and the degree of similarity with their own relationships. Such family based revenge has already been reported in monkeys (Aureli et al. 1992; Judge 1982; Seyfarth & Cheney, in press). This suggests that even the cognitive processes underlying vendettas are not unique to human beings.

Personality, self-control, and welfare-tradeoff ratios in revenge and forgiveness

doi:10.1017/S0140525X12000349

Daniel Balliet and Tila M. Pronk

Social and Organizational Psychology Department, School of Psychology and Education, VU University, Amsterdam, 1081 BT Amsterdam, The Netherlands.

dp.balliet@psy.vu.nl t.pronk@vu.nl

http://www.psy.vu.nl/nl/over-de-faculteit/medewerkers-alfabetisch/medewerkers-a-b/d-balliet/index.asp

http://www.psy.vu.nl/nl/over-de-faculteit/medewerkers-alfabetisch/medewerkers-m-p/pronk-t/index.asp

Abstract: We address how trait self-control and trait concern for others relate to the concepts of monitored and intrinsic Welfare Tradeoff Ratios (WTRs), respectively, and how recent work on personality, revenge, and forgiveness are informed by the adaptationist perspective proposed in the target article. We also discuss how the proposed adaptationist perspective provides clues to some previously puzzling findings on revenge.

In the target article, McCullough et al. forward a timely adaptationist framework for conceptualizing revenge and forgiveness. According to this theoretical perspective, revenge and forgiveness are evolved psychological mechanisms that regulate the interpersonal behavior of a victim in response to the harm-doing of a transgressor. Central to their model is the mechanism of Welfare Tradeoff Ratios (WTRs) in conducting cost–benefit analyses of interdependent social behaviors. WTRs are psychological mechanisms that compute the relative welfare of a target's welfare compared to one's own welfare. In the target article the authors make a distinction between intrinsic and monitored WTRs. Intrinsic WTRs involve welfare tradeoffs that consider the indirect benefits a particular relationship has for the individual's own welfare, whereas monitored WTRs involve computations of a welfare-tradeoff taking into consideration the target's ability to monitor and respond to one's behavior. In light of this theoretical framework, we review recent research relating personality with revenge and forgiveness and suggest how this perspective may explain some recent unexpected findings in the literature.

Research suggests that forgiveness can be challenging, and that forgiveness is facilitated by the use of self-control – an ability to monitor and regulate behavior to achieve long-term goals (Balliet 2010; Pronk et al. 2010). Additionally, prior research has found that self-control increases the ability to positively weigh others' outcomes (i.e., WTRs) during interdependent social interactions (Balliet & Joireman 2010). Thus, self-control is not only

relevant for revenge/forgiveness, but may also affect the proximate mechanisms outlined in the target article, namely WTRs. Can the relation between self-control, forgiveness, and revenge be understood by the adaptationist model? Here we suggest that self-control may work together with evolved psychological mechanisms (e.g., WTRs) to affect revenge and forgiveness.

Specifically, individual differences in self-control may aid our understanding of *monitored* WTRs. People who monitor their behavior in relation to goals during social interactions, thereby exerting self-control, may be more thoughtful about how their own and other's current behavior may affect future outcomes. Another possibility is that self-control is used to compare other's perceived WTRs to one's own WTR, and this process may affect revenge and forgiveness. Although previous research suggests that self-control may enable people to inhibit their desire for revenge in order to maintain valuable social relations, as we discuss below, self-control can also enable people to become more vengeful. Other traits may relate to individual differences in the calibration of *intrinsic* WTRs (such as social value orientations; see Balliet et al. 2009). Importantly, both features of personality that affect monitored and intrinsic WTRs may interact to affect revenge and forgiveness.

Recently, Balliet et al. (2011a) measured intentions of revenge in response to a partner's initial transgression during an iterated prisoner's dilemma (and maximizing difference game). They found that trait self-control negatively related to revenge in response to their partner's defection, but only amongst individuals who were less concerned for others' outcomes, relative to their own outcomes (i.e., low intrinsic WTRs). In this experimental context, participants were thought to be interacting for several trials of the dilemma. One implication of this finding is that self-control may affect calculating concern for anonymous others, and especially in situations when another has an ability to respond and punish one's behavior. Certainly, in the context of each iterated game, mutual cooperation is in the long-term self-interest for both parties. Thus, self-control may be a general ability that works by adjusting (monitored) WTRs to manage social relations and achieve long-term outcomes for the individual. A second implication is that the effect of self-control on revenge may depend on a person's intrinsic WTR.

An unexpected finding in recent work is that positive intrinsic WTRs can lead to stronger revenge motivation in response to a perceived transgression, but only when people have the time and exert self-control to think about the costs and benefits of revenge (Balliet et al. 2011a; Perunovic & Holmes 2008). Perhaps high intrinsic WTRs establish expectations of social behavior that are easily violated, and self-control may initiate a comparison between own and an other's perceived WTRs that may subsequently encourage revenge in an attempt to get the other to recalibrate their monitored WTR to reach an equilibrium with their own. Yet, for individuals who have a low intrinsic WTR, the use of self-control may result in attempts to display an increase in their own WTR toward the other. Interestingly, in both accounts self-control may encourage strategies to reach an equilibrium between one's own and the other's WTR.

A second finding not easily explained by existing theories is that punishment more effectively increases cooperation when punishments are costly to deliver (Balliet et al. 2011b). Prior theorizing suggests that reduced costs of punishment make punishment more effective at promoting cooperation. Yet, this finding may indicate the importance of others' perceived WTRs for revenge and forgiveness. Not only do people possess their own WTR, but also cognitive mechanisms disposed toward understanding others' WTRs, and these hold important implications for both own intrinsic *and* monitored WTRs. Perhaps costly punishments communicate that punishment is delivered out of concern for the relationship or group and so may be more effective by simultaneously increasing the transgressor's own intrinsic WTR as well as the monitored WTR.

As we reflected on the ability of an adaptationist perspective to guide research on revenge and forgiveness, we noticed in several instances that this perspective could be meaningfully related to conclusions from our own research and is able provide clues to some previously puzzling findings. Specifically, trait self-control and trait concern for others may affect forgiveness by the calibration of monitored and intrinsic WTRs, respectively – a possibility that deserves future research attention. Overall, we are excited about the possibility of this model directing future research. Managing social relations certainly provided an abundance of challenges in our ancestral environment that were directly relevant to survival and reproduction. Taking revenge to deter harm and forgiving others to maintain vital social relations are likely two important ingredients that have enabled humans to successfully navigate the social environment.

Pathways to abnormal revenge and forgiveness

doi:10.1017/S0140525X12000507

Pat Barclay

Department of Psychology, University of Guelph, Guelph, Ontario N1G 2W1, Canada.

barclayp@uoguelph.ca

http://www.uoguelph.ca/nacs/page.cfm?id=229

Abstract: The target article's important point is easily misunderstood to claim that all revenge is adaptive. Revenge and forgiveness can overstretch (or understretch) the bounds of utility due to misperceptions, minimization of costly errors, a breakdown within our evolved revenge systems, or natural genetic and developmental variation. Together, these factors can compound to produce highly abnormal instances of revenge and forgiveness.

In the target article, McCullough et al. do an admirable job of arguing that revenge is not a disease, and instead may be an adaption to prevent exploitation. This approach is long overdue in many social sciences, as it moves away from pejorative preconceptions about behaviours we don't like. As with any adaptive explanation for behaviour, there is a high risk of the authors' argument being misunderstood to claim that all instances of revenge should be adaptive. Such misunderstandings regularly occur with other evolutionary explanations of human social behaviour. As such, the authors' argument requires extension to examine when revenge and forgiveness will overstretch (or understretch) the bounds of utility, and why.

A complete explanation of revenge and forgiveness will include errors of absence as well as errors of excess. While there are popularized cases of ridiculous revenge, we often overlook the excessive "lack of revenge" or excessive forgiveness. These are the things that fill books like *Chicken Soup for the Soul*. They are potentially equally maladaptive, but we don't see them as "errors" because we "like" this behaviour (see also Wakefield 1992). There is likely an optimal level of revenge and forgiveness for any situation. Too little revenge is an insufficient deterrent, but too much revenge invites further retaliation (Barclay 2008). Too little forgiveness prevents reparation of a relationship, but too much forgiveness invites future exploitation (Axelrod 1984). Finding the optimal level of revenge involves "brinksmanship" (Daly & Wilson 1988), a difficult game when people have imperfect information about the world or about others' past and future intentions (e.g., Todd 2001). Because of such constraints, no evolved psychological mechanism is expected to produce optimal results in every single instance, but is expected to be adaptive on average (Haselton & Buss 2000; Nesse 2005; Barclay 2011). The following are some causes of excessive or insufficient

revenge or forgiveness, and when combined in one individual, they could result in markedly abnormal behaviour indeed.

Misperceptions of costs and benefits. To produce adaptive levels of revenge and forgiveness, our brains must use environmental cues of the costs and benefits. Assessing these costs and benefits is no small task, as one must assess one's strength relative to a transgressor (and possibly an audience), audience presence and characteristics, one's need, the risk, the relationship value, and a host of other factors. Naturally, there is error associated with assessing any of these (Barclay 2008). Sometimes these errors will balance out, but when they don't, they will compound to produce more revenge or forgiveness than is optimal. Statistically speaking, these misperceptions alone will cause deviations from optimality that are normally distributed about the optimum, with most individuals near optimality but with some individuals displaying excessively high or low revenge or forgiveness.

Misperceptions of others' actions. The optimal level of revenge and forgiveness likely depends on others' intentions, both in terms of the reasons for their past actions and their intended future behaviour. Assessing such intentions is a mind-reading game, and is also prone to error. "Rules of thumb" based on past experience will only sometimes be right, and will sometimes overestimate hostility. Based on this, we might predict that people who are better at reading others' intentions will produce more optimal levels of revenge and forgiveness.

Minimizing costly errors. Different errors have different costs, and natural selection has presumably designed our emotions so that we avoid committing more costly or more frequent errors (Haselton & Buss 2000; Nesse 2005). If being too vengeful is more costly than being insufficiently vengeful, then our revenge systems should be biased towards producing less revenge than is "needed," and vice versa. A similar argument holds for forgiveness. Which is more costly, excessive or insufficient revenge (or forgiveness)? This is probably an empirical question. In fact, the answer may vary in different social environments depending on the frequency and importance of exploitation (bias towards excessive retaliation) and long-term cooperation (bias towards excessive forgiveness). By focusing on the costs and frequencies of these different errors, we can predict when we will observe excessive vengeance or excessive forgiveness.

Genetic or development noise variation. The target article outlines a number of tasks performed by our revenge and forgiveness systems, each of which involves many steps. As with any complex trait, each of these sub-tasks will be affected by multiple genes and environmental influences. Because these influences can combine in different combinations, it will cause natural variation about an optimum for each sub-task, resulting in some individuals in the tail ends of excessive revenge.

Pathologies within the revenge systems. Although the target article suggests that revenge is not a "disease," it does leave open the possibility of genuine diseases *within our evolved revenge systems*. Some individuals might indeed have something "broken" in the brain areas responsible for assessing costs, benefits, and intentions, or for producing an appropriate level of revenge. For example, if a (subconscious) assessment of costs tends to inhibit revenge, then anything that damages the brain's inhibitory systems will prevent this inhibition and will result in excessive revenge. Also, if someone is insensitive to costs or punishment in general (e.g., psychopaths), then there will be nothing to lower their vengefulness down to optimal levels. In other words, the capacity for revenge is not pathological, nor is the acting on that capacity, but there can be pathologies associated with expressing that capacity. It is these pathologies that probably produce the types of revenge and forgiveness that make newspaper headlines (Barclay 2008).

These are but some of the potential causes of abnormal levels of revenge and forgiveness. Some will result in normally distributed variation in revenge and forgiveness, whereas others will cause systematic biases towards excess (e.g., pathologies, error management). Altogether, they show how not every instance of revenge will be beneficial, nor will every instance of forgiveness. Thus, we can extend the framework that McCullough et al. provide to make predictions about "abnormal" levels of revenge and forgiveness.

The cultural shaping of revenge

doi:10.1017/S0140525X12000350

Stephen Beckerman

Anthropology Department, Pennsylvania State University, University Park, PA 16802.

stv@psu.edu

Abstract: There are interesting parallelisms between McCullough et al.'s article and studies of revenge presented by French legal anthropologist Raymond Verdier, particularly as regards the discussion of the increasing likelihood of revenge with increasing social distance. Additionally, the observation that many peoples speak of revenge in the language of debt and repayment, links it with exchanges of benefits as well as costs.

Substantial parts of this interesting target article by three psychologists, McCullough, Kurzban, and Tabak (McCullough et al.), are strikingly congruent with studies of revenge conducted by anthropologists over the last three decades. Among the most important publications in this arena are the four volumes edited by French legal anthropologist Raymond Verdier and his collaborators (Verdier 1981a; 1986; Verdier & Poly 1984; cf. Courtois 1984). In his introduction to that work, Verdier (1981b) points out that in many (perhaps most) societies vengeance is spoken of in terms of debt and repayment, the vocabulary in which people talk about the owing and paying of goods and services – and most importantly, the same terms in which the exchange of women as brides is discussed. Indeed, one of the most common means of terminating an actual or potential blood feud is for the family or lineage of the killer to turn over one of its daughters as a wife to the family or lineage of the homicide victim, the woman's life-giving capacity being taken as compensation for the life than was taken. This perspective potentially amplifies the applicability of the Welfare Tradeoff Ratio (WTR) to include all fitness relevant exchanges, comprehending benefits as well as costs in a single calculation.

Another area of convergence arises from the authors' remark that "we expect revenge to be less likely in the context of kin, people with whom one has an ongoing exchange relationship (…), friends and allies (…), and long term mates" (target article, sect. 4.1, para. 3). Verdier (1981b; 2008) distinguishes three increasingly distant categories of social relations – *identity*, *adversity* (by which he means that the actors on the poles of the relationship are adversaries, but not usually permanent enemies), and *hostility* – each marked by a characteristic form of retribution. (These categories map rather well to the three spheres of reciprocity – generalized, balanced, and negative – proposed by Sahlins (1972) to classify the varieties of material exchange.) Within identity, the first and closest category of social relations (e.g., the family, the clan), violent revenge is forbidden. To kill or injure someone in that tight circle would only be to compound the initial injury to oneself. Retribution is characteristically left to the workings of supernatural forces.

It is in the second category, adversity, (e.g., different clans within the same tribe, neighboring tribes that intermarry) that the cultural elaboration of revenge flourishes, often with elaborate rules stipulating what constitutes an injury calling for revenge, who ought (or must) take revenge, who is eligible and who is ineligible as a target, and where and when and how it is permissible to take revenge, and what sort of revenge is mandated. The typical goal in this realm of adversity is to achieve a balance of injuries,

after which peaceful relations and their mutual benefits are resumed.

In the most distant category, that of hostility, there are no ongoing beneficial relations between the groups involved (e.g., strangers, invaders, etc.), no attempts at a balance of injuries, and no ameliorating rules. The goal is to crush, if not exterminate, the enemy who committed the initial injury.

From this viewpoint, it is perhaps inaccurate to expect that "revenge is more frequently used in societies in which social institutions for settling grievances are generally viewed as weak" (target article, sect. 3.2.2, para. 3). A more ethnographic approach might assert that revenge, in the sphere of adversity, is a means of settling grievances – and further, that revenge, in addition to its individual purpose of raising one's WTR with respect to a previous adversary, has the social purpose of restoring peaceful relations between adversaries. Perhaps the best example of this function of revenge is found in Exodus 21:23. "A life for a life, an eye for an eye, a tooth for a tooth." The explicit point of this passage is that once the proportional act of revenge has taken place, the exchange is to be considered closed and no further hostilities are permitted.

On the evolutionary origins of revenge and forgiveness: A converging systems hypothesis

doi:10.1017/S0140525X12000362

Richard J. Crisp and Rose Meleady

Centre for the Study of Group Processes, School of Psychology, University of Kent, Canterbury, Kent CT2 7NP, United Kingdom.

r.crisp@kent.ac.uk

rm379@kent.ac.uk

www.kent.ac.uk/psychology/people/crispr/

www.kent.ac.uk/psychology/people/meleadyr/

Abstract: McCullough et al. argue that humans possess evolved computational systems for implementing retaliatory behavior that both deters aggression and promotes subsequent reconciliation. However, they do not apply this analysis to the sphere of intergroup relations. We believe their model can be usefully extended to this domain and discuss why this would be possible, pertinent, and productive.

In their target article, McCullough et al. propose that cognitive systems for revenge and forgiveness result from psychological adaptations that helped solve certain social problems encountered through the course of human evolution. While the authors develop an important new perspective on the evolutionary origins of such behaviors, they make several assumptions that, we believe, limit the scope and applicability of their own thesis. Most notably, they "take no position on whether the psychology that governs the operation of revenge systems (…) also evolve to regulate behavior in intergroup contexts" (sect. 2.2.4, para. 2). In what follows we argue that the authors' model should be extended to include intergroup contexts, and that it correspondingly has potentially important implications for understanding the nature of conflict and cooperation between groups.

Research into the origins of group behavior has established the key role that intergroup relations have played in the evolution of social cognition. This work suggests that "the group" constitutes the "mind's natural environment," acting as the interface between an individual and their physical environment (Caporael 1997). As such fitness should be correlated with the development of functionalized cognitive mechanisms supporting the development of relationships that necessarily include intergroup comparisons (Caporael & Baron 1997). Being able to effectively make such comparisons is critical, because in ancestral environments clear cognitive representations of ingroup (us) versus outgroup (them) boundaries affords significant functional benefits such as maximizing security and limiting the potential for disease transmission.

If we accept that intergroup cognition has an evolutionary origin, could the same basic system be used in intragroup contexts (and vice versa)? McCullough et al. suggest not, and that a different cognitive architecture may have evolved to govern intergroup conflict compared to intragroup conflict (sect. 2.2.4, para. 2). Our contention is that, in evolutionary terms, systems for regulating intra- and intergroup behavior should be intimately linked because they involve precisely the same computational requirements. In other words, determining whether to retaliate to a member of one's coalition, and subsequent reconciliatory behavior, can apply just as much (and perhaps more crucially) when negotiating competitive and cooperative intergroup relations.

Evidence for this "converging systems hypothesis" can be found in research on intergroup relations that has demonstrated the operation of precisely the same mechanisms as revealed in studies of *intra*group public goods dilemmas (including some of the same studies cited by the authors; e.g., Fehr & Gächter 2002; Kollock 1998; Ostrom 1990; Yamagishi 1986). Specifically, converging systems are evident when it comes to general incentives for cooperation and non-cooperation that apply to larger social contexts in which groups compete for resources (i.e., intergroup contexts). Take Sherif's (1966) classic realistic group conflict theory. This intergroup theory is based on the assumption, shared by much work on intragroup public dilemmas, that conflicts are "rational" in the sense that opposing groups have incompatible goals and compete for scarce resources (Taylor & Moghaddam 1987). Bornstein (2003) similarly argues that intragroup dilemmas are embedded within, and indeed characterize, intergroup conflict and cooperation. He notes that before it is rational for groups to compete, it must be rational for the individual group members to do so. In other words, the benefits associated with the outcome of intergroup conflict (e.g., security, territory, political power, status, pride) are public goods which extend to all group members regardless of their individual contribution. As Dawes (1980) states, "Soldiers who fight in a large battle can reasonably conclude that no matter what their comrades do they personally are better off taking no chances; yet if no one takes chances, the result will be a rout and slaughter worse for all the soldiers" (p. 170). In sum, the types of social dilemmas cited by the authors are inherently embedded within, and indeed characterize, situations of intergroup conflict and cooperation. This suggests that common processes can determine both intra- and intergroup outcomes, and provides a basis for predicting a similar level of convergence when it comes to the computational systems involved.

The potential relevance of this convergence becomes apparent when considering the central role revenge and forgiveness have played in contemporary theories of intergroup conflict and its reduction. In fact, when McCullough et al. conclude that "the attention placed on revenge and forgiveness in the psychological literature has been somewhat limited and undertheorized" (sect. 7, para. 2), we would point out that revenge and forgiveness have been key concepts in over 60 years of research on the potential for intergroup contact to reduce conflict (Allport 1954; Brown & Hewstone 2005; Pettigrew & Tropp 2006). Mechanisms underlying processes of forgiveness and reconciliation are particularly central to this work, and relevant cognitions and emotions such as perspective-taking and empathy are well-specified (Hewstone et al. 2006; Paolini et al. 2004; Turner et al. 2008). Furthermore, anthropological and sociological studies have shown that contact outcomes, such as the formation of relationships that cut across tribal boundaries, promote forgiveness and reconciliation (Coser 1956; Deutsch 1973; Evans-Pritchard 1940; LeVine & Campbell 1972). This work suggests that groups have an evolved propensity to engage in intergroup contact precisely because it can give rise to cross-cutting affiliations that ensure stability and security (i.e., it

is more difficult to have conflictual relations with a group based on territory that is simultaneously an ally according to common ancestry; Crisp & Hewstone 2007). Notably, recent work has begun to specify the computational mechanisms through which these positive intergroup relations can be established (Crisp & Turner 2009; 2011). Considering how these contributory cognitive systems evolved will help us develop a clearer understanding of how and when contact can be successful in tackling contemporary intergroup conflicts. It may also help us move closer to understanding how these computational systems enabled ancestral coalition building, and through this, the construction of complex societies.

An eye for an eye: Reciprocity and the calibration of redress

doi:10.1017/S0140525X12000519

Andrew Dellis[a] and David Spurrett[b]

[a]*Research Unit in Behavioural Economics and Neuroeconomics, Department of Economics, and Brain and Behaviour Initiative, Department of Psychiatry and Mental Health, University of Cape Town, Groote Schuur Hospital, Observatory, 7935, Cape Town, South Africa;* [b]*Department of Philosophy, University of KwaZulu-Natal, Durban 4041, South Africa.*

andrew.dellis@uct.ac.za

Spurrett@ukzn.ac.za

Abstract: General systems for reciprocity explain the same phenomena as the target article's proposed revenge system, and can explain other cooperative phenomena. We need more reason to hypothesise a specific revenge system. In addition, the proposed calculus of revenge is less sensitive to absolute magnitudes of revenge than it should be.

We have two related critical points concerning the target article. The first is that the evidence surveyed is consistent with a simpler and more general hypothesis. This is clearer in the case of revenge, where the relevant hypothesis is that there is adaptation for reciprocity, which applies to both welfare increasing and welfare reducing actions by others. The second is that the welfare trade-off ratio (WTR) hypothesised to determine acceptable harm/benefit ratios between individuals fails to account for the importance of absolute magnitude of revenge.

The literature in behavioural ecology pays significant attention to reciprocity of cooperative and non-cooperative behaviours, including reciprocation of cheating and aggression with withholding of cooperation (Trivers 1971). Conversely, it pays little or no attention to revenge as a separate topic. Welfare increasing actions, such as grooming, infant access, and resource sharing, are no less suitable candidates for reciprocation than welfare decreasing ones, such as aggressive assault and preventing resource access. So, for example, we see dynamic and context sensitive interactions between rates of welfare-decreasing and -increasing behaviours among female baboons, including aggression against low-ranking ones, and grooming of high-ranking by low-ranking ones (Barrett et al. 2002). A relatively unified system is needed for trading off expected costs and benefits of available actions (including actions classifiable as altruistic, antagonistic, reconciliatory, or punitive). This is so even if its implementation is distributed and if some sub-functions – such as cheat detection – are specialised. Hypothesising a single general reciprocity system (or reciprocity sensitivity) instead of a specific revenge system means not having to hypothesise a separate system for cooperation and altruism. In addition, hypothesising a single system is consistent with a large body of established work on the unified neural representation of rewarding and aversive outcomes (Montague & Berns 2002).

The best way to make a case *against* a general reciprocity system and in favour of a specific revenge adaptation would be to provide evidence of dissociation. For example, if in some cases of local brain damage, genetic intervention, and so on, there were individuals who could reciprocate altruism but not exploitation or attack, or vice versa, the hypothesis of specific revenge adaptations would be on stronger ground. There could be other considerations favouring a specific revenge adaptation that has features not predicted by a general reciprocity disposition, but the target article does not specify them or give evidence that they are satisfied.

It may seem as though a general reciprocity system would struggle to account for forgiveness. But the evidence regarding forgiveness is equivocal. Among nonhuman primates reconciliation behaviours following conflict have been documented in many species, and are sometimes taken to serve the function of repairing valuable relationships. These behaviours are not universal (see, e.g., Kappeler [1993] on the absence thereof among ring-tailed lemurs). Among the species in which they are observed, they exhibit properties that are partly at odds with the hypothesised forgiveness among humans. In particular, unlike forgiveness, which is a victim-initiated process, reconciliations are typically initiated by former aggressors (e.g., post-conflict grunting by former aggressor baboons – see Castles & Whiten 1998a). In addition the same outcomes, in respect of reduced stress by the victim of aggression, and reduced rates of subsequent attack from the previous aggressor, are observed in cases where instead of reconciliation there is redirected aggression by the former victim to an individual lower in the hierarchy (Aureli & van Schaik 1991a). The target article is silent on why redirected aggression might be a substitute for reconciliation.

Our second concern is that the schematic quantitative proposal regarding revenge is, in at least one important respect, incomplete. The basic proposal in the target article is that agents maintain and update welfare tradeoff ratios (WTRs) for other agents, which determine acceptable harm/benefit outcomes. The problem with simple ratios is that they are unable to account for a key feature of revenge or punishment, at least when described by many recorded cultures, which is that it is often quantified in absolute terms. Ratios on the other hand merely fix a range of outcomes, including ones where the revenge is both much larger and much smaller than whatever provoked it. (Put in monetary terms, a hostile WTR of 1 to 10 between me and some other individual says I'd willingly pay a dollar to see them lose ten dollars, or two to see them lose twenty, or a million to see them lose ten million, or…). Absolute specification is much more common. So, various ancient legal systems of which we have surviving records prescribe lists of penalties for recognised offences, which vary in type (death, corporal punishment, specific fines) but are all absolute in the sense of requiring a single death, corporal penalty, or a fine of some magnitude rather than a ratio consistent with a large range of outcomes. Examples include the Babylonian Code of Hammurabi from the 18th century BC, and the older Sumerian Codes of Lipit-Eshtar and Ur-Nammu, as well as more recent but still ancient codes such as the Hittite Laws. At various places in the Old Testament a principle of extracting equal injury in retaliation (at least for some crimes with human victims) is stated, including the specific formulation of "eye for eye" (Exodus 21:23–25; Leviticus 24:19–21; Deuteronomy 19:20–21). When Lady Capulet says "Romeo slew Tybalt, Romeo must not live" (*Romeo and Juliet*, Act 3, Scene 1), she expresses the view that utterly independent of the energetic cost and inconvenience, a penalty of a specific magnitude must be exacted.

An important critical response here would argue that these codified systems represent attempts to combat the ongoing damage that could ensue in the absence of a way of concluding some conflicts, and that they may even be attempts to solve a problem arising from a simpler WTR system for revenge. Were that the case one might expect that closely related social primates lacking legal systems would sometimes engage in runaway reciprocity and revenge governed by welfare tradeoff ratios and insensitive to absolute magnitudes.

Towards a multifaceted understanding of revenge and forgiveness

doi:10.1017/S0140525X12000374

Ramzi Fatfouta,[a] Arthur Jacobs,[a,b] and Angela Merkl[a,c]

[a]*Dahlem Institute for Neuroimaging of Emotion (D.I.N.E.), Cluster of Excellence "Languages of Emotion", Freie Universität, Berlin, D-14195 Berlin, Germany;* [b]*Department of Educational Sciences and Psychology, Freie Universität, Berlin, D-14195 Berlin, Germany;* [c]*Department of Psychiatry, Charité-University Medicine Berlin, Campus Benjamin Franklin, D-14050 Berlin, Germany.*

ramzi.fatfouta@fu-berlin.de **http://www.languages-of-emotion.de/en/**

ajacobs@zedat.fu-berlin.de

http://www.languages-of-emotion.de/en/dine/

angela.merkl@charite.de **http://psychiatrie.charite.de/en/**

Abstract: We focus on two aspects: First, we argue that it is necessary to include implicit forgiveness as an additional adaptive behavioral option to the perception of interpersonal transgressions. Second, we present one possible way to investigate the cognitive-affective underpinnings of revenge and forgiveness: a functional MRI (fMRI) approach aiming at integrating forgiveness and revenge mechanisms into a single paradigm.

Mere acceptance or implicit forgiveness? McCullough et al. propose "acceptance" as a means of tolerating the transgressor's Welfare Tradeoff Ratio (WTR; Tooby et al. 2008). Despite establishing one plausible behavioral option to the infliction of interpersonal harm, we deem McCullough et al.'s notion elusive for two reasons:

First, it neglects the possibility that accepting an injustice may be equivalent to what Exline and Baumeister (2000) termed *implicit forgiveness*. Whereas explicit forgiveness (a) directly relates to an existing debt and (b) acknowledges one's willingness to absolve the transgressor from his or her guilt (e.g., "I forgive you for cheating on me"), the offended person may also implicitly express forgiveness – either by downplaying the transgression, for example, by saying "It's okay," or wordlessly by maintaining contact with the transgressor and thus not ceasing the relationship. Along this line, the concept of implicit forgiveness differs from mere acceptance in the sense that a motivation to engage in retaliatory aggression changes into forgiving motivations, but no direct reference is made to the cancelled debt (cf. Exline & Baumeister 2000). We therefore think that implicit forgiveness may also represent a possible adaptive mechanism when perceiving a social norm violation.

Second, one important question remains unanswered: What if the victim, for reasons of introspective limits (Nisbett & Wilson 1977), *explicitly* reports that he or she is okay with the aggressor's WTR, yet, *implicitly*, it reveals that he or she is still holding grudges? In order to provide empirical evidence for McCullough et al.'s suggestion that acceptance is marked by an absence of residual motivation to engage in retaliation, the use of indirect forgiveness measures seems beneficial. Consistent with the notion that cognitive accessibility of the self through introspection is limited (Greenwald & Banaji 1995; Greenwald et al. 2002), indirect measures such as the Implicit Association Test (IAT; Greenwald et al. 1998) capture information processing that resides outside conscious control or awareness (e.g., Asendorpf et al. 2002; Greenwald & Banaji 1995; Wilson et al. 2000). Hence, developing an indirect forgiveness measure could not only favor a multi-method approach (Eid & Diener 2006), but also elucidate phenomenological differences between acceptance and implicit forgiveness.

Investigating the neural substrates of revenge and forgiveness: An fMRI approach. We concur with McCullough et al. that making adaptive decisions about revenge and forgiveness encompasses multiple cognitive components. Judging the forgivability of another's actions is one such component (Farrow et al. 2001). For example, a post-therapy fMRI study with patients suffering from posttraumatic stress disorder (PTSD) revealed altered forgivability judgments following symptom resolution, suggesting that traumatic experience changes brain responses to social cognition tasks (Farrow et al. 2005). Hayashi et al. (2010) further observed in healthy subjects that the ventromedial prefrontal cortices (vmPFC) play a central role in forgivability judgments for moral transgressions regarding deceptive behavior.

Although game theory has clearly advanced our understanding of social decision-making (Sanfey 2007), deciding upon forgiveness versus revenge is neither an all-or-nothing proposition (Hayashi et al. 2010), nor a purely rational or computational process. Unfortunately, most psychological and neuroscientific studies have largely tried to assess forgiveness via hypothetical moral scenarios (e.g., "Which of the following crimes you would see as more forgivable?"; Farrow et al. 2001, p. 2434) instead of via real and personally relevant stimuli. Notably, all forgiveness scenarios have been based on judging actions of unknown individuals (cf. Farrow et al. 2001). It thus remains unclear whether activations in the proposed brain regions are specific to forgiveness or whether they mirror social and/or moral judgment more globally (Hayashi et al. 2010).

Here, we would like to propose that an alternative, non–game-theoretic approach is also well suited for investigating the roles of both revenge and forgiveness. More specifically, we hold an autobiographical memory fMRI paradigm (for a review, see Maguire 2001) as particularly promising. The main benefit of using events with personal significance is that by reliving memories from the participants' own history, forgiveness- and revenge-related emotions (e.g., anger, fear) that were connected to those memories are subjectively re-experienced (Rubin 2005; Svoboda et al. 2006). An ancillary advantage is that this procedure allows for the induction of a personally relevant affective state (Wagner et al. 2011) and this, in turn, may activate brain regions in a similar manner as the original emotional event (Buchanan 2007). We conjecture that this approach is able (a) to induce an intrapsychic instance of real-life forgiveness and revenge, and (b) to elicit affect-laden memory relevant for forgiveness and revenge.

We further believe that this approach could shed light on McCullough et al.'s pronounced suggestion that the evolved forgiveness systems "inhibit revenge" (sect. 4.2, para. 2). If so, emotional reliving of forgiveness events should recruit prefrontal brain areas robustly associated with cognitive control and emotion regulation. This proposal stands in line with recent empirical evidence, demonstrating that executive functioning (i.e., cognitive control processes) is negatively associated with retaliatory aggression, and that this effect is mediated by lower levels of revenge motivation (i.e., higher forgiveness) (Wilkowski et al. 2010). Relatedly, a series of studies by Pronk and colleagues have illustrated that cognitive control is positively associated with forgiveness (see Pronk et al. 2010). Pronk et al. (2010, p. 128 f.), and even directly allude to the possibility that the prefrontal cortex (involving the anterior cingulate cortex [ACC]) may be implicated in the down-regulation of negative affect.

To conclude, from a neuroscientific perspective, it is not only important to explore which specific brain areas are involved in revenge and forgiveness processes, but also to demonstrate (a) the interplay of both cognitive *and* affective signature(s) within these processes; (b) the involvement of emotionally arousing memory in these processes; and (c) its advantage over game-theoretic approaches to the study of forgiveness and revenge. Using McCullough et al.'s framework as a starting point, these important issues open up fields of interesting research questions to be addressed in the future.

An implausible model and evolutionary explanation of the revenge motive

doi:10.1017/S0140525X12000386

Herbert Gintis

Santa Fe Institute, Santa Fe, NM 87501, and Department of Economics, Central European University, Nádor utca 9, 1051 Budapest, Hungary.

hgintis@comcast.net

http://www.santafe.edu **http://people.umass.edu/gintis**

Abstract: McCullough et al.'s target article is a psychological version of the reputation models pioneered by biologist Robert Trivers (1971) and economist Robert Frank (1988). The authors, like Trivers and Frank, offer an implausible explanation of the fact that revenge is common even when there are no possible reputational effects. I sketch a more plausible model based on recent research.

The target article by McCullough et al. is a psychological version of reputation models pioneered by biologist Robert Trivers (1971), who called revenge "moralistic aggression" and the associated emotion "moral outrage," and economist Robert Frank (1988), who called it "passion within reason." This model of reputation effects, with similar assumptions, was deepened in the economics literature (Fudenberg et al. 1994; Kreps & Wilson 1982; Schmidt 1993).

The psychological dimension to the reputational model enters because the proximate motives for seeking revenge in human societies are generally not to enhance reputation, but rather to obtain satisfaction by harming the offender. Moreover, revenge is common in humans even when its cost is greater than its expected reputational gains, a fact that is difficult to reconcile with the biological and economic reputational models. One example is strong reciprocity which, in a social dilemma context where there is no opportunity for reputation formation, involves being predisposed to cooperate initially and as long as others reciprocate, and to punish non-contributors at personal cost (Bowles & Gintis 2011; Fehr & Gachter 1998; Gintis 2000; Gintis et al. 2005). Another example is third party punishment where, even under conditions of anonymity, an individual punishes an agent who has harmed a stranger, or who has committed a social norm violation that does not affect the punisher (Buchholtz et al. 2008; Fehr & Fischbacher 2004).

McCullough et al. explain the psychology of revenge and its widespread occurrence in situations where deterrence is not involved by arguing that in our hunter-gatherer prehistory, revenge had positive fitness effects by establishing the reputation of the revenge-seeker as an individual who is not to be exploited and who will defend his family and allies. A genetically based human cognitive deterrence system thereby became adaptive. This deterrence system persists in modern life where it is maladaptive because, by contrast with the Pleistocene, contemporary social conditions include many one-shot and anonymous interactions. The absence of one-shot and anonymous interactions in the human hunter-gatherer societies of the Pleistocene explains why evolution gave rise to a cognitive deterrence system that does not condition revenge behavior on the level of expected future returns.

There are three problems with this argument. First, modern humans routinely distinguish between situations in which reputation building is possible and situations in which it is not, and cooperate much more in the former case (Bowles & Gintis 2011, Ch. 3). Assuming that this capacity is a cognitive adaptation, there must have been frequent and fitness-relevant non–reputation-building interactions in our evolutionary history.

Second, even in a world of repeated interactions among well-acquainted individuals, anonymous interactions (e.g., hiding a kill from others) are common in contemporary hunter-gatherer societies (Kaplan et al. 1984; Hawkes 1993; D. S. Wilson 1998), and hence doubtless in Pleistocene and early Holocene times as well. Indeed, such behavior is routinely recorded in chimpanzees, and hence is likely an attribute of our most recent common ancestor some eight million years ago (Boehm 2011; de Waal 1997).

Finally, it is not the case that general individuals in prehistoric hunter-gatherer communities were life-long social interactants. The evidence supporting this assertion comes from Late Pleistocene climate records, archaeological records of the causes of death, and genetic evidence bearing on exogamy and migration. Neither the likely size of groups, nor the degree of genetic relatedness within groups, nor the typical demography of foraging bands is favorable to the view that Late Pleistocene human groups sustained cooperation through either kin-based or reciprocal altruism. Rather our ancestors were cosmopolitan, civic-minded, and warlike. They benefited from far-flung coinsurance, trading, mating, and other social networks, as well as from coalitions and, if successful, warfare with other groups (Bowles & Gintis 2011, Ch. 6).

I offer the following sketch of an alternative model of revenge-seeking, which treats this motive as a form of moral behavior: Individuals seek revenge not when they have been hurt, but when they have been morally wronged, or when they countenance others violating the social norms of the community.

Like other organisms, humans have a preference ordering over states of affairs, and they act to best achieve their desired states of affairs, given the material and informational resources available to them. These preferences are strongly influenced by genetic predispositions, but they are affected by group culture. Culture for humans is not merely a set of learned techniques, but also a set of moral values that are internalized by group members (Parsons 1967). The capacity to internalize values through socialization is an evolved adaptation of humans (Durkheim 1902/1967; Gintis 2003; Simon 1990), and accounts both for cultural diversity across societies and (limited) cultural stability across generations.

Human social cooperation is governed not by genes alone, but by social norms that are widely embraced by social actors, and act as moral values present in individual preference functions. We term these *other-regarding preferences* (Gintis 2009). Individuals incorporate moral values in their actions by trading off the costly attainment of other-regarding goals against self-regarding goals. The ubiquity of altruistic cooperation and altruistic punishment around the world suggests that these values are strong genetic predispositions, although the evidence indicates that their expression is strongly modulated by local cultural values (Henrich et al. 2004; 2005; 2006). There are several plausible models of the evolution of these predispositions (Bowles & Gintis 2011; Boyd et al. 2010; Gintis 2000). There are also plausible evolutionary models of the internalization of norms, the mechanism by which moral values become represented in the individual's preference ordering (Boehm 2011; Bowles & Gintis 2011; Gintis 2003).

In this alternative framework, revenge and forgiveness can be self-regarding acts aimed at deterring malefactors and warning others of the cost of aggression. But revenge can also be an other-regarding act carried out to redress wrongs on a purely moral level. This explains why individuals punish not only those who hurt themselves, their families, and their allies, but more generally those who violate social norms. It also explains why individuals will seek vengeance against aggressors even when there is no possible deterrent effect.

Revenge without redundancy: Functional outcomes do not require discrete adaptations for vengeance or forgiveness

doi:10.1017/S0140525X12000520

Colin Holbrook, Daniel M. T. Fessler, and Matthew M. Gervais
Center for Behavior, Evolution, and Culture and Department of Anthropology, University of California, Los Angeles, Los Angeles, CA 90095-1553.
cholbrook01@ucla.edu dfessler@anthro.ucla.edu
mgervais@ucla.edu
http://cholbrook01.bol.ucla.edu/
http://www.sscnet.ucla.edu/anthro/faculty/fessler/
http://www.anthro.ucla.edu/people/grad-pages?lid=4411

Abstract: We question whether the postulated revenge and forgiveness systems constitute true adaptations. Revenge and forgiveness are the products of multiple motivational systems and capacities, many of which did not exclusively evolve to support deterrence. Anger is more aptly construed as an adaptation that organizes independent mechanisms to deter transgressors than as the mediator of a distinct revenge adaptation.

Following Sell et al. (2009), we agree with McCullough et al. that multiple factors shape responses to being wronged (e.g., whether

the transgressor is a close ally, kin, or someone likely to exact high costs due to a status or formidability differential), and that this process is intimately related to the motivational profile of anger. McCullough et al. go further, however, by apparently proposing the existence of additional specialized psychological adaptations to enable deterrence. It is most parsimonious to attribute the deterrence-related computations reviewed by the authors to the emotion "anger," operating in conjunction with (1) mechanisms that transcend the domain of interpersonal conflict (e.g., norm-acquisition, future forecasting, perspective-taking), (2) an attitudinal system that regulates a wide variety of behaviors, and (3) systems related to other motivations, such as reputation management.

Consider the complex case of indirect deterrence. In our view, the computational demands described by McCullough et al. in this regard are met by evolved capacities to categorize events, assume others' perspectives, forecast the future, and weigh costs against benefits. These capacities are directed and organized over short time spans by the emotion of anger (Fessler 2010; Tooby & Cosmides 2005), and over longer time spans by the more enduring attitude of hatred, an evaluative representation that tracks and reacts to the fortunes of an other whose principal relationship with the self is as a source of costs inflicted in zero-sum contexts (Gervais & Fessler, under review). Hence, on the one hand, if by "an evolved cognitive system that implements … deterrence" (target article, Abstract) the authors mean a functionally specialized system that evolved expressly for this purpose, then we would argue that redundant algorithms for deterrence-related event categorization, perspective-taking, cost–benefit analysis, and so on, seem implausible – why engineer new content-dedicated devices when a bricolage of existing devices will satisfy? On the other hand, if the authors concede that there is no uniquely bounded "revenge adaptation," but contend that, nonetheless, the outputs of this bricolage can be treated *as if* they are produced by such an adaptation, given that they address a unified domain (i.e., "revenge" is a recurrent adaptive task), then we would argue that the authors have mistaken a folk category (cost infliction motivated by anger and hatred following transgression) for a nonexistent natural kind. There are many kinds of deterrence that do not stem from the anger-hatred nexus (e.g., swatting a dog in order to teach it not to steal food off the table), and hence neither constitute "revenge" in any ordinary sense of the word, nor involve the core motivational components of the bricolage at issue.

The above critique holds for each of the observations adduced by McCullough et al. As further evidence of special design, the authors discuss strategic calibrations made in light of culturally and individually varying exigencies, such as whether the putative adaptation operates in a legalistic society that punishes retaliatory violence, or in a weak soma likely to be injured in combat. We agree that humans adaptively modulate deterrence behavior in light of social and personal contexts, but, again, see no reason to postulate specialized subroutines of a revenge adaptation. Cultural norm acquisition mechanisms (Sripada & Stich 2007) are sufficient to enable learning of locally accepted ways of resolving conflict. Reputation management mechanisms are also implicated, moderating retributive behavior to the extent that the reputational consequences of how one responds to transgression vary, with some societies valorizing, and others demonizing, violent retribution (Fessler 2006). This suggests only the interaction of distinct psychological motives (i.e., to punish, to protect one's reputation, etc.), not, as the authors imply (sect. 3.1.2, paras. 1–4), that the supposed vengeance system contains a customized reputation circuit. This explains why the presence of onlookers can magnify not only violence, but also charitable giving (Harbaugh 1998) and shame displays (Fessler 2004) – reputation management systems operate in tandem with, and may potentiate or vitiate, other systems.

As evidence of a forgiveness adaptation, McCullough et al. observe that transgressors' relatedness, past friendship, or opportunity to injuriously counterattack, mitigate the severity of deterrent responses to transgressions. The competing perspective that we have applied to the revenge adaptation applies here as well. Although humans likely do take fitness-relevant factors such as relatedness, prior cooperation, and relative status/formidability into account during conflicts, it is more parsimonious to ascribe these calibrations to the operation of other systems (e.g., for affiliation in the case of transgressive friends or kin, or fear in the case of formidable adversaries) that moderate anger than to propose new, highly redundant pathways engineered to facilitate strategic détente.

We have argued that the postulated wholes (adaptations for revenge and forgiveness) are not greater than the sums of their parts (perspective-taking, event categorization, norm-acquisition, future forecasting, reputation management, etc.). The proposed adaptations do not appear to possess domain-specific content beyond components that, although useful in calculating deterrence, mostly evolved for other reasons. Anger is indeed considered to have evolved to deter harmful transgressors by inflicting costs or withholding benefits, and has demonstrated unambiguous domain-specificity in this regard (e.g., Fessler & Gervais 2010; Lazarus 1991; Sell et al. 2009). McCullough et al. characterize anger as the proximal mediator of the proposed revenge adaptation, but this appears to needlessly multiply entities. The crux of the issue is whether a vengeance adaptation evolved with specialized mechanisms to compute factors such as the likelihood, type, and severity of reprisals, the intentions of the transgressor, social consequences, status differentials between self and transgressor, prior history of cooperation with transgressor, kinship with transgressor, and so forth, or whether these diverse variables are taken into account through the simultaneous operation of multiple domain-specific modules operating within the same mind, perhaps coordinated by anger in the short term, and hatred in the long term. In both scenarios, retaliatory behavior is moderated by personal, cultural, and situational factors; adjudicating the issue is therefore a problem of theory rather than of missing or disputed data. Given these options, we advocate the latter alternative because it is simpler, kludgier, and therefore more evolutionarily plausible.

Revenge and forgiveness or betrayal blindness?

doi:10.1017/S0140525X12000398

Sasha Johnson-Freyd[a] and Jennifer J. Freyd[b]

[a]*Department of Human Evolutionary Biology, Harvard University, Peabody Museum, Cambridge, MA 02138;* [b]*Department of Psychology, University of Oregon, Eugene, OR 97403.*

johnsonfreyd@college.harvard.edu
https://sites.google.com/site/johnsonfreyd/
jjf@uoregon.edu
http://dynamic.uoregon.edu

Abstract: McCullough et al. hypothesize that evolution has selected mechanisms for revenge to deter harms and for forgiveness to preserve valuable relationships. However, in highly dependent relationships, the more adaptive course of action may be to remain unaware of the initial harm rather than risk alienating a needed other. We present a testable model of possible victim responses to interrelational harm.

In the target article, McCullough et al. offer the intriguing hypothesis that mechanisms for revenge in humans have evolved to deter harms and that forgiveness mechanisms evolved to compensate for the possibility or consequences of revenge in order to preserve valuable relationships. They refer to four possible responses to interrelational harm: acceptance, forgiveness, avoidance, or revenge. Such responses, however, are

contingent on the victim *perceiving* the harm, yet such awareness is not always apparent or adaptive. Extrapolating from Betrayal Trauma Theory (Freyd 1996), we suggest a different way to structure these concepts (see Fig. 1), where their "avoidance" and "acceptance" are included in our *withdrawal* and *unawareness*, respectively. True acceptance requires awareness; however, in many cases (we argue in *most* cases), what looks like acceptance to an outside observer is actually motivated unawareness.

If a victim is *aware* of the harm, he or she then has the choice to *demand repair*, *withdraw* from the relationship, *forgive* the perpetrator, or enact *revenge* (Fig. 1). After a demand for repair or withdrawal, the victim's next options depend on the perpetrator's response. If the response is a good one, *reconciliation* might occur, whereas if the response is negative, it constitutes a new harm and the suite of behavioral options re-starts.

Importantly, the option of awareness depends upon the victim's degree of empowerment in the interpersonal relationship in which the harm occurred. As the target article notes, a victim's response depends heavily on his/her relationship with the perpetrator. For example, McCullogh et al. predict that relationships with expected future value are more likely to be forgiving. However, categories of interpersonal relationships involve more than just their perceived future value.

Dependence is a particularly important dimension of relationships. Being dependent on others for material and emotional support has profound implications for adaptive responses to harm. Betrayal Trauma Theory (Freyd 1996; DePrince et al. 2012) posits that when a victim is significantly dependent on the perpetrator, it may be adaptive to remain unaware of the harm the perpetrator imposed. A dependent victim is essentially required to maintain the relationship with his or her aggressor. Most of the options shown in Figure 1 that follow *awareness* may be detrimental to the relationship on which the victim depends and therefore are not adaptive.

Betrayal blindness is theorized to be a basic response among humans. Empirical research suggests that betrayal blindness is both common and psychologically important for the victim (DePrince et al. 2012; Freyd et al. 2007). It is likely that betrayal blindness has played an important role in human evolution: For humans to survive into adulthood, they had to live through periods of significant dependence (such as childhood). Dependence continues in various forms (e.g., due to illness or resource asymmetries) throughout the lifespan. Furthermore, although there is variation in severity, harm in interpersonal relationships is ubiquitous. Thus, every individual who reproduced successfully maintained important interpersonal relationships with people who had more power than them and sometimes caused harm. Selection pressure may have created evolutionarily ancient human victims who had the ability to remain unaware of interrelational harm.

Why would a person remain unaware rather than acknowledge and either ignore (pretend not to see) or "accept" a betrayal? We propose that such pretending is often not adaptive because of the resources necessary for maintenance and the risks associated with failure. If the victim is very young (infant or toddler), he or she may not have the cognitive capacity to pretend and thus be required to remain unaware in order to preserve the relationship. Even in adulthood, most humans may find it difficult to be effective pretenders. For example, in trying to feign happiness with a perpetrator, a victim may have trouble smiling in a seemingly authentic way (i.e., Duchenne smiling; see Ekman & O'Sullivan 2006). There is great risk to being a poor pretender: losing a necessary (or apparently necessary) relationship. Even when effective pretending is possible, it may be very costly to cognitive capacity by consuming attention resources that would then not be available for other tasks. It is hard to see how such a risky and resource-demanding process (feigning unawareness/acceptance) could be adaptive.

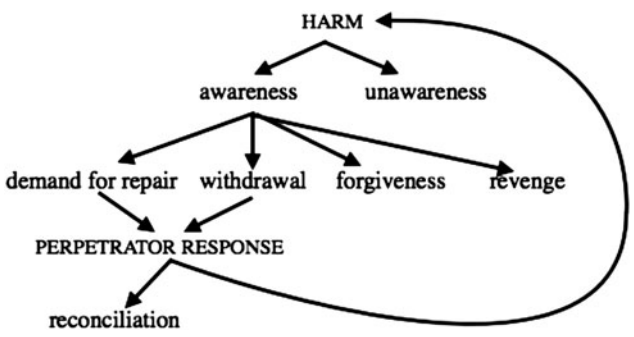

Figure 1 (Johnson-Freyd & Freyd). Responses to interrelational harm.

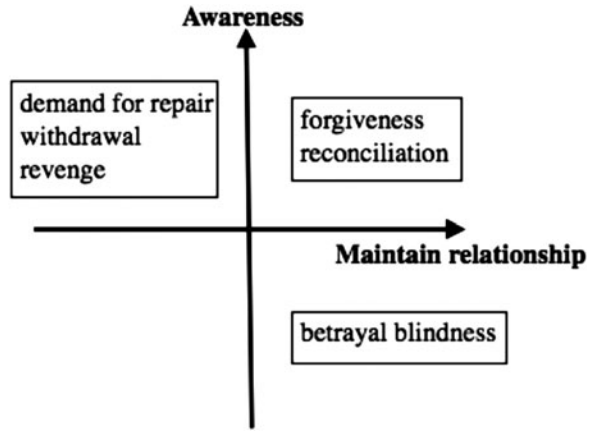

Figure 2 (Johnson-Freyd & Freyd). Proposed dimensions of responses to interrelational harm.

McCullogh et al.'s description of behavioral options (sect. 4.4) fails to give significant attention to the variation in awareness that distinguishes the possible responses. For instance, the authors' concept of "acceptance" may actually be better understood as *unawareness* (betrayal blindness). In other words, a victim may appear to "accept" a harm by remaining unaware of it. In contrast, both revenge and forgiveness constitute explicit actions in response to interrelational harm that necessitate explicit thought and understanding about that harm and the interpersonal relationship between the victim and the aggressor.

We can understand different behavioral responses to harm by organizing them on two orthogonal axes: (1) degree of awareness, and (2) whether the victim wants to maintain the relationship (Fig. 2). For example, a victim may *forgive* an aggressor when he or she wants to maintain the relationship and is highly aware, whereas a victim may remain *blind* to the betrayal when he or she wants to maintain the relationship with the aggressor and thus is *unaware* of the harm. In this model, forgiveness may be most common when the victim holds significant power in the relationship. Betrayal blindness is predicted to be frequent when the perpetrator holds significant power. A question awaiting future research is how tightly connected harm awareness is with empowerment.

Another interesting research question concerns the evolution of the awareness necessary for various behavioral responses to harm. Forgiveness and revenge seem *behaviorally* similar to other responses (e.g., reconciliation and counter-aggression) but *psychologically* different because of the difference in cognitive awareness. Do nonhuman animals exhibit the responses of revenge and forgiveness? Such comparative research might help us further understand the evolution of the different possible responses to interrelational harm.

ACKNOWLEDGMENTS

We thank Katie Hinde and Steve Pinker for their helpful feedback on a prior draft of this commentary.

It takes more to forgive: The role of executive control

doi:10.1017/S0140525X12000532

Johan C. Karremans and Reine C. van der Wal

Behavioural Science Institute, Radboud University Nijmegen, 6500 HE Nijmegen, The Netherlands.

j.karremans@psych.ru.nl **R.vanderwal@psych.ru.nl**

http://www.ru.nl/socialpsychology/faculty/dr_johan_karremans/

http://www.ru.nl/socialpsychology/phd-students/reine-van-der-wal/

Abstract: The target article's evolutionary approach provides an excellent framework for understanding *when* and *why* people retaliate or forgive. We argue that recent findings on the basic processes in forgiveness – particularly, the role of executive control – can further refine the authors' proposed model. Specifically, the lack of executive control may restrict the explanatory power of relationship value and exploitation risk.

The adaptationist analysis of revenge and forgiveness as offered by McCullough et al. provides a very welcome overarching theoretical framework to better understand these concepts. As the authors note, empirical research on revenge and forgiveness has been rather scattered, and mainly driven by mini-theories. The authors have done an impressive job to integrate a host of previously isolated findings to support their evolutionary approach to forgiveness and revenge. Their analysis leads to several interesting testable predictions about when people will be more, or less, likely to take revenge, and when they are more likely to forgive a transgressor.

The target article concludes with proposing a computational model that helps humans to decide whether to take revenge or forgive an offender, or essentially, which response offers the most fitness benefits. Ultimately this decision depends on perceived relationship value (as indexed by psychological constructs such as closeness and commitment) and perceived future exploitation risk, and their interaction. As cited by the authors, research with both human and nonhuman subjects has revealed strong support for the relationship value prediction (e.g., Finkel et al. 2002; Karremans & Aarts 2007; Watts 2006). Recently, in line with their evolutionary argument, we have demonstrated that the positive association between interpersonal closeness and forgiveness is robust across several different (both independent and interdependent) cultures – albeit with some cross-cultural variation regarding the strength of the association (see Karremans et al. 2011).

However, although the theorized computational system offers a very useful tool for understanding *when* and *why* humans forgive or take revenge, less attention is paid to the *how* of revenge and forgiveness. Recent studies have provided important insights into the basic processes that lead to forgiveness, demonstrating that executive functioning – in particular the cognitive ability to control and inhibit impulsive responses (as assessed with Stroop-like measures) – is an important facilitator of forgiveness (e.g., Pronk et al. 2010; Wilkowski et al. 2010; cf. Finkel & Campbell 2001). Whereas the initial impulsive response to a transgression often is to retaliate, individual differences in executive control are positively associated with the ability to inhibit such retaliatory and negative affective responses, and instead to respond in a forgiving manner (Pronk et al. 2010).

Importantly, it appears that individuals low in executive control have difficulty forgiving an offending relationship partner even when the partner is someone they feel close to – or, to use McCullough et al.'s terminology, *even* when relationship value is high. In a recent series of studies in primary schools, we have found initial evidence that 11- and 12-year old children are more likely to forgive their friends than non-friends, but crucially, that this "relationship value" effect was more strongly pronounced among children high in executive control (Van der Wal et al. 2012). In fact, although based on the relationship value hypothesis we should have expected stronger forgiveness when the transgressor is a friend rather than non-friend across the range, children low in executive control basically did not show this effect. Similar effects were found in a study with late adolescents, revealing that closeness only predicted forgiveness among participants high in executive control. These findings suggest that high relationship value generally leads to the recruitment of executive control in order to down-regulate negative emotions toward the offender, *unless* the individual lacks such executive control resources. Thus, relationship value is not always the best possible predictor of forgiveness – at least not for everyone, or under all circumstances (e.g., when executive control resources are temporarily depleted).

Admittedly, this literature has so far not looked at how executive control may be related to perceived exploitation risk. A possible prediction based on the authors' proposed model is that low executive control individuals may have more difficulty in estimating exploitation risk, which may prevent them from forgiving valuable relationship partners. Or, alternatively, low executive control might disrupt the entire computational process, failing to successfully integrate relationship value and exploitation risk information in order to decide whether or not to forgive.

The fact that low executive control hinders forgiveness, even in the face of high relationship value (and possibly, low exploitation risk), raises intriguing and complex questions. For example, and following the authors' adaptationist logic, do individuals low in executive control – which is strongly genetically determined (see Friedman et al. 2008) – adopt alternative strategies to minimize the fitness costs of their relative inability to forgive valuable relationship partners? Or, as with the example of sex differences provided in the target article, does executive control modify the costs and benefits of forgiving valuable others, such that the lack of forgiveness may be less detrimental to the valuable relationships of individuals low versus high in executive control? Do low executive control individuals in some way compensate for the loss of fitness benefits from their relative struggle to forgive? Although very speculative, perhaps low executive control individuals – or more broadly, any individual with a lower forgiveness propensity for whatever reason – may seek out relationship partners who possess particularly well developed conflict-resolution skills.

To conclude, whereas in the target article the authors have built an evolutionary theoretical account of forgiveness and revenge by integrating largely dispersed research findings, in turn this account inspires many novel and specific questions. However, although relationship value and exploitation risk are the factors that help to explain when and why revenge or forgiveness are the most adaptive and thus most likely responses, they may be limited in addressing how revenge and forgiveness actually occurs. Yet, knowledge about the basic processes that describe *how* forgiveness occurs also informs us about *when* forgiveness or revenge is the most likely response. Hence, we believe that recent findings on executive control and forgiveness – and more generally, any past and future findings on the role of the proximate cognitive and neural mechanism involved in forgiveness and revenge – can help to further inform and refine the authors' theoretical approach.

Revenge: Behavioral and emotional consequences

doi:10.1017/S0140525X12000404

Vladimir J. Konečni

Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0109.

vkonecni@ucsd.edu

http://psychology.ucsd.edu/people/profiles/vkonecni.html

Abstract: This commentary discusses dozens of ecologically powerful social-psychological experiments from the1960s and 1970s, which are

highly relevant especially for predicting the *consequences* of revenge. McCullough et al. omitted this work – perhaps because of its misclassification as "catharsis" research. The findings are readily accommodated by Konečni's anger-aggression bidirectional-causation (AABC) model and can be usefully incorporated in an adaptationist view of revenge.

It is commendable that the authors of this excellent adaptationist account of an important aspect of human social interaction are concerned that "for some crucial questions about the revenge and forgiveness systems […] data are scant" (McCullough et al., target article, sect. 7, para. 2). However, this is not entirely accurate. The main objective of this commentary is to discuss some very relevant experimental work on revenge that has apparently escaped the attention of McCullough et al., in the hope that these neglected findings and the associated theoretical ideas can be usefully incorporated into their broad view.

Unlike the majority of findings cited by McCullough et al., the work in question is not from the domain of economic games, which is significant given the external-validity doubts that can be raised about games research with regard to the genuineness of participants' motivation and, especially, emotion. Instead, the data come from social-psychological behavioral experiments published in the1960s and 1970s (in top-tier journals), in which ecologically powerful procedures were used that the subsequent human-research regulations made difficult to implement. Furthermore, some of these experiments dealt with issues that may arise in long-term human dyadic relationships; such data may contribute to the authors' complex analysis of repeated "effective updating" (sect. 3.1.1, para. 2).

The key questions are these: What are the behavioral and emotional *consequences* of revenge? How might these outcomes influence both the avenger's (AV) and the initial offender's (IO) computations of the present and future costs and welfare tradeoff ratios (WTRs)?

Most of the data come from a three-stage research paradigm: (1) IO's offense against AV (such as insults); (2) AV's behavioral retaliation against IO (such as fictitious electric shocks); and (3) obtaining dependent measures of AV's arousal, anger, and *additional* behavioral aggression against IO. These experimental results are informative about the short- and longer-term, both internal (sympathetic arousal, rated anger) and external (additional aggressive behavior), consequences for AV (and for IO as the target of any additional aggression) of the retaliatory actions previously executed by AV against IO.

To summarize the data which have been obtained *as a function of revenge*:

1. A sharply *reduced* amount of *immediate* (*additional* – that is, post-revenge) *aggression by AV against IO* (and also against substitute or "scapegoat" targets) – not only in comparison with the behavior of would-be avengers who did not have a prior opportunity for retaliation (Doob & Wood 1972; Konečni & Doob 1972; Konečni & Ebbesen 1976), but also of those who were required to perform tasks (math problems) that minimized the likelihood of anger-producing rumination (Konečni 1975a). In fact, even observing the IO (allegedly) in pain (Bramel et al. 1968) or (allegedly) hurt by someone else (Doob & Wood 1972) decreased the amount of retaliatory aggression directed by the offended person at the culprit.

2. A significantly *decreased level of AV's physiological arousal* (that had been sharply raised by IO) compared to various control groups (Hokanson & Burgess 1962; Hokanson et al. 1963; Hokanson & Shetler 1961). Revenge decreases physiological arousal quickly. More generally, because aggressive responses apparently succeed in terminating noxious stimulation emanating from others more effectively than other responses, *ceteris paribus*, they acquire arousal-reducing properties (Konečni 1975a; Patterson & Cobb 1971).

3. Auxiliary findings that are theoretically congruent with those in point (2) have also been obtained: As a function of behavioral revenge against IO, avengers display a restored affinity for complex stimulation (Konečni et al. 1976) and a reduced level of alcohol consumption (Marlatt et al. 1975).

4. A significantly *lower level of AV's self-rated anger*, compared to participants without a retaliatory opportunity, but, importantly, *as high a level of AV's dislike for IO* as that observed in appropriate control participants (Konečni 1975a; Konečni & Doob 1972).

The entire observed pattern of findings, (1) to (4), can be accommodated by Konečni's (1975a; 1984) anger-aggression bidirectional-causation model (AABC). The model also predicts, because of the data in the above-mentioned points (2) and (4), that the *future* execution of aggressive acts by AV against IO would be more likely in long-term dyads (and occur sooner in the offense-revenge sequence): The original angry, righteous avenger may become an anger-free ("cold-blooded") bully who strikes with little or no provocation. Such pre-emption complicates the computation of long-term WTRs beyond what McCullough et al. have proposed for revenge, possibly with large errors along the long road of adjustment or even a complete breakdown of the relationship (often with dire consequences). *Retaliatory pre-emption* – an unprovoked attack camouflaged as retaliation for an (imaginary) offense – is also relevant for the computation of "indirect [third-party] deterrence" (sect. 3.1.2).

Another important fact – predicted by the AABC model – that should influence the computations by both AV and IO is that the *amount* of revenge is strongly affected by the random arousal-related circumstances in which the initial offense occurs. Specifically, the amount of revenge has been observed in experiments to increase as a function of additional (*unrelated*) stressors that are present concurrently with, or immediately following the initial offense. When AVs do strenuous physical exercise (Zillmann et al. 1972) or listen to loud and complex tones (Konečni 1975b), their retaliation against IOs is more severe than that performed by controls. Therefore, from both AV's and IO's computational perspective, the context of the initial offense is important – as is the perceived intentionality of both the offending and vengeful actions.

The research described above has been largely ignored – for various (bad) reasons. It was pigeonholed as "catharsis" and falsely related to the outmoded "hydraulic" model of Freud and Lorenz, or to Aristotle's "pity and terror" – but, significantly, not to Plato's correct judgment of the benefits of revenge. There was the dubious idea that watching boxing films, fantasy aggression, or children attacking inanimate targets (none of these genuine *vengeful* activities) should reduce aggression – yet the opposite, and correct, result is predicted by the AABC model. A slew of inadequate experimental procedures has been used to disprove straw "catharsis" hypotheses and reach the socio-politically desirable conclusion that "aggression breeds aggression" (something easily achieved, according to AABC). Fortunately, sound evolutionary thinking (in the target article) has finally imposed a reality check on wishful thinking.

The fuzzy reality of perceived harms

doi:10.1017/S0140525X12000416

Sara Konrath[a] and Irene Cheung[b]

[a]*Research Center for Group Dynamics, Institute for Social Research, Ann Arbor, MI 48104;* [b]*Department of Psychology, The University of Western Ontario, London, Ontario N6A 5C2, Canada.*

skonrath@umich.edu **icheung4@uwo.ca**
www.sarakonrath.com **http://www.wix.com/icheung4/uwo**

Abstract: We review two subjective (mis)perceptions that influence revenge and forgiveness systems. Individual differences predict more (e.g., narcissism) or less (e.g., empathy) revenge, with the opposite pattern for forgiveness. Moreover, differences in victim versus perpetrator perceptions can influence revenge and forgiveness systems, perpetuating never-ending cycles of revenge. These two examples point to the need for

theories of revenge and forgiveness to address the role of cognitive and motivational biases in the functionality of such behavioral responses.

When it comes to revenge and forgiveness, there is no black and white world where harms are objective. Perceptions matter, whether the misperceptions of individuals who overestimate or underestimate their deservingness of benefits, or misperceptions that stem from the fuzzy nature of "who started it." Such misperceptions can exaggerate harm, and ultimately lead to miscalibrated revenge responses relative to initiating circumstances. Theories of revenge and forgiveness must account for cognitive and motivational processes that serve to inflate or reduce perceptions of harm.

First, what happens when individuals consistently miscalibrate their estimations of others' welfare tradeoff ratios (WTRs) toward the self? Although McCullough et al. touch on the role of individual differences, they mainly focus on ones related to physical strength (e.g., sex), which directly maps onto one's ability to enact revenge. However, individual differences in the propensity toward revenge and forgiveness cannot all be explained this way.

For example, it is likely that people scoring high on the personality trait narcissism overestimate others' WTRs toward themselves, and if so, they would perceive continual violations of these expected WTRs. This would lead to over-active revenge systems to try to increase others' regard for their welfare. Practically, this would manifest itself as increased sensitivity to others' harms to the self, over-reactive anger responses, and a lower likelihood of forgiveness, each of which are correlates of narcissism (Exline et al. 2004; McCullough et al. 2003; Rhodewalt & Morf 1998). Although in past research males often scored higher than females in narcissism, such sex differences are small and are becoming smaller over time (Twenge et al. 2008). And most research on narcissistic anger and aggression finds that these effects occur independently of sex (Twenge & Campbell 2003). Thus, narcissists should be likely to see themselves as deserving of unquestioning respect, and to (mis)perceive violations of their expected WTRs, regardless of sex. This rules out the possibility that such individual differences are only explained by the power to successfully enact revenge.

Similarly, people high in dispositional empathy may chronically miscalibrate their WTRs in the opposite direction, and have under-active revenge systems and over-active forgiveness systems (Macaskill et al. 2002; Stuckless & Goranson 1992). This could make these individuals ripe for potential exploitation, leaving open questions about the evolution of such individual differences. Again, although there are sex differences in self-reported empathy, these differences disappear in physiological measures (Eisenberg & Lennon 1983; Lennon & Eisenberg 1989). Thus, it is unlikely that empathy is associated with less revenge and more forgiveness because empathic individuals are less able to successfully enact revenge. A number of other personality variables are also consistently associated with more or less revenge and forgiveness (Mullet et al. 2005).

One way to explain such individual differences in revenge and forgiveness may be to consider the role of interdependence (see sect. 4.2, para. 2). For example, those scoring high in narcissism see themselves as less interconnected and interdependent with others (Konrath et al. 2009), and do not place a high value on relationships (Foster et al. 2006). Thus, they may not be concerned about the relational costs of enacting revenge for even minor perceived transgressions. Because they are always on the lookout for new and better relationship partners (Campbell & Foster 2002; Campbell et al. 2002), the potential to lose current partners might not bother them too much. However, even if this were the case, it would only explain their *individual* motivations for being overly vengeful, and not the *evolutionary* function – unless this type of behavior offered them some sort of survival or reproductive advantage.

Other misperceptions are also important to consider. For example, how do differences in victim versus perpetrator perceptions influence revenge and forgiveness systems? Victims and perpetrators do not always see eye-to-eye on the impact of harms, such that victims perceive harms as having continuing implications for their relationships, whereas perpetrators perceive harms as being isolated incidents without long-lasting implications for their relationships (Baumeister et al. 1990; Zechmeister & Romero 2002). Given such discrepancies in perceptions of harms, victims may retaliate against perpetrators to deter future harms, but these actions may in turn be seen as overreactions or unjust by initial perpetrators, which can ironically lead to feelings of victimization in them. Thus, the roles of the victim and perpetrator can easily reverse and perpetuate cycles of revenge (Schumann & Ross 2010; Stillwell et al. 2008). In other words, when both parties' perceptions of the harms are not calibrated, revenge cycles may be initiated.

McCullough et al. touch on counter-revenge as a cost to revenge and the "echo effect," but more elaboration is needed. When victims and perpetrators are in revenge cycles, how do these cycles end if their actions are driven by (mis)perceptions? What triggers forgiveness in these cycles? Or, at what point do relationships simply dissolve? Also, what is the evolutionary function of revenge cycles?

Victims may seek revenge to change perpetrators' WTRs toward them. However, because perpetrators may also see themselves as victims, they may also try to increase avengers' WTRs toward them. Thus, both victim and perpetrators may feel compelled to increase their retaliation level in order to change WTRs, which can cause irreparable damages to relationships and make it surprising that forgiveness ever occurs at all. Perhaps one function of revenge cycles is to give individuals an opportunity to assess the value of their relationships, so that they can withdraw from potentially unproductive ones (Kearns & Fincham 2005). That being said, it is also possible that revenge cycles may be more likely to occur *after* individuals have already decided to dissolve a relationship. In other words, such misperceptions might be more common in the presence of unproductive relationships, and may serve as a catalyst toward dissolution.

We have reviewed two subjective (mis)perceptions that may influence revenge and forgiveness systems, pointing to the need for theories of revenge and forgiveness to address the role of cognitive and motivational biases in the functionality of such behavioral responses.

On the differential mediating role of emotions in revenge and reconciliation

doi:10.1017/S0140525X12000428

David Leiser and Lisa Joskowicz-Jabloner

Department of Psychology, Ben-Gurion University, Beer Sheva 84105, Israel.
dleiser@bgu.ac.il
joskolis@gmail.com **http://www.bgu.ac.il/~dleiser/**

Abstract: McCullough et al. suggest that revenge and forgiveness rest upon risk computation. Risk computation is implemented by emotions that evolved for additional functions, giving rise to phenomena such as betrayal aversion and taboo-tradeoffs, and specific patterns of forgiveness we have documented. A complete account of revenge and reconciliation should incorporate broader constructs from social psychology, including emotions and values hierarchies.

McCullough et al. analyze revenge and reconciliation in terms of computation of risk and the welfare tradeoff ratio (WTR). Their functional analysis does not clarify how these mechanisms are implemented, but suggests that the same computational psychological mechanisms are always involved. This leaves out a significant component of a complete account.

Risk computation cannot explain why most of us are willing to forgo benefits, including fitness gains, to avoid the experience of betrayal. People will make different decisions on whether to take a risk, depending on whether the potential loss will be inflicted by a person or by some probabilistic machine with the same probability (for "betrayal aversion," see Bohnet & Zeckhauser 2004; Koehler & Gershoff 2003). The reason for this is that a perceived risk of betrayal triggers emotions which, in turn, influence decision making. Moreover, attitudes to trust and the impact of fear of betrayal are culture-dependent, indicating something more than a case by case cost–benefit analysis. In some cultures, betrayal is so aversive that essentially no risks are taken, and people will forgo significant potential gains attainable with substantial probability, in order to not incur the risk of betrayal (Bohnet et al. 2010).

McCullough et al. analyze the different costs and benefits involved in exacting revenge from close as opposed to distant associates. Our work (Joskowicz-Jabloner & Leiser, forthcoming) demonstrates that trust violations evoke different patterns of emotions, depending on the closeness of the perpetrator. These emotions in turn predict avenues for reconciliation, as well as, crucially, the ineffectiveness of certain remedies to promote reconciliation. We established the existence of (at least) two distinct betrayal-domains that differ in the pattern of emotional response evoked, in the actions that relieve negative emotions and the influence of values held by the victim on the emotional response. In the social norms domain, betrayal evokes predominantly anger-related emotions that can be alleviated effectively, whereas in the personal domain more profound negative emotions are elicited by betrayal and fewer actions can relieve them.

For example, according to the analysis of McCullough et al., offers of compensation or hyper-compensation should be effective. In our study, participants did perceive monetary (hyper-) compensation as highly effective when the perpetrator was a stranger, but as an inappropriate remedy when the perpetrator was a friend. In fact, such offers to a friend compound the injury and make forgiveness more difficult – a situation described as a "taboo trade-off." A taboo trade-off (Fiske & Tetlock 1997) is an incommensurable comparison between two values, in which one value is desecrated by being weighed against another, for example to attach "a monetary value to one's friendship" (p. 256). Because taboo trade-offs themselves trigger negative emotional, behavioral and cognitive responses, offering compensation for a betrayal in the personal domain is likely to be perceived by the betrayed as an additional act of betrayal. In terms of appraisal theory, the victim of personal betrayal is forced to undertake an ongoing assessment of the situation (Ellsworth & Scherer 2003). The betrayal victim's reappraisal of the relationship is first based on the betrayal incident and then by the offer of monetary compensation, which implies that the perpetrator seeks or sees an instrumental relation to the victim.

The reason for this finding is quite obvious: In the case of friends and close allies, the computations about WTR, past investment, and future expectations are mediated by the feelings or emotion of *personal* trust, that are absent with respect to strangers. Moreover, emotions such as personal trust or honor are of use beyond computation of WTR, on a cumulative or case by case basis, and relate to broader psychological structures, such as values. The value hierarchies espoused by people modulate the functioning of such emotions. McCullough et al. would of course not dispute this, but the consequences need to be spelled out. Let us take an example from our work: We found that the experience of trust-betrayal is influenced by the importance of the values "Power" (indicating a concern for self-interest) and "Benevolence" (expressing the concern for the welfare of others) (Schwartz 2005) for the betrayal victim. Power amplifies the emotional response and also enhances the effectiveness of several relief actions, whereas benevolence may ease negative emotions, in particular emotions of anger and regret, but it strengthens the efficiency only of apology. Similarly, Stouten et al. (2005; 2006) found that in the face of norm violators, "pro-socials" (the benevolence-oriented; see De Cremer & Van Lange 2001) are more concerned with fairness and conflict resolution, whereas "proselfs" (the power-oriented) focus on efficiency.

In view of these considerations, a complete account of revenge and forgiveness should go beyond WTR and incorporate several central constructs from social psychology. To explain specific behavioral patterns, it is necessary to identify the specific emotions that are used to tally and compute the relevant risks and benefits in each case. These emotions evolved for a variety of broad functions, out of which WTR computation is but one. They are differentially evoked and involved according to the circumstances, and related to yet broader psychological structures, such as values and values hierarchies.

No such thing as genuine forgiveness?

doi:10.1017/S0140525X12000544

Mark G. McCoy and Todd K. Shackelford
Department of Psychology, Oakland University, Rochester, MI 48309.
mgmccoy@oakland.edu shackelf@oakland.edu
www.ToddKShackelford.com

Abstract: McCullough et al. propose adaptations that motivate forgiveness when the potential benefits of continuing the relationship outweigh the costs incurred by the transgression. The costs incurred are definite, whereas future benefits of forgiveness are only probabilistic. This situation exposes the forgiver to cheating in the form of repeat transgression. Adaptations motivating genuine forgiveness are therefore unlikely to evolve.

McCullough et al. present a convincing case for the evolution of psychological mechanisms that motivate revenge. They define revenge as harm inflicted on someone for a perceived transgression. According to the authors, revenge can (1) deter the transgressor from inflicting future harms, (2) deter others from inflicting future harms, or (3) deter others from withholding expected benefits.

In this commentary, we address McCullough et al.'s proposal that genuine forgiveness, like revenge, is produced by evolved mechanisms. According to the authors, forgiveness is motivated by the desire to maintain a relationship with the transgressor. They argue that exacting revenge can produce costs for the vengeful that are greater than the potential benefits enjoyed from forgiveness and subsequent continuation of the relationship with the transgressor. According to McCullough et al., therefore, functional assessments of the potential benefits of forgiveness as contrasted with the costs of revenge are key design features of the evolved mechanisms that produce forgiveness.

The costs inflicted by the initial transgression are definite whereas the benefits gained from forgiveness are only probabilistic. The potential future benefits of forgiveness may occasionally outweigh the costs to the vengeful of exacting revenge. That is, the relationship with a forgiven transgressor may not yield future benefits, and the opportunity to exact revenge may have passed. Forgiveness therefore will have produced a net cost for the forgiver.

Moreover, the forgiven transgressor may inflict further costs on the forgiver, especially given the recent history of violating with impunity the forgiver. Furthermore, the now multiply-transgressed forgiver risks earning a reputation among the local group as someone who can be violated with impunity – inviting transgressions from others. Psychological adaptations that motivate genuine forgiveness therefore may be unlikely to evolve. What is more likely to evolve are adaptations that motivate *feigned* forgiveness, lulling the transgressor into reconciliation with the potential to exact revenge when he least expects it. Feigned forgiveness also would allow one to evaluate the potential future benefits of

maintaining the relationship with the transgressor. If the benefits of maintaining the relationship exceed the benefits of exacting revenge, feigned forgiveness can produce net benefits. Feigned forgiveness also provides for the opportunity to maintain cautious surveillance of the transgressor, reaping any benefits of continuing the relationship while remaining alert to attempts at repeat transgression. A feigned forgiver "forgives but does not forget."

For clarity, we contrast briefly a genuine forgiver with a feigned forgiver. Although the genuine forgiver has sincerely reconciled with and is preparing to maintain a future relationship with the transgressor, he is open to future violations by the transgressor and from others knowledgeable of the transgressor's successful violations of the genuine forgiver. The genuine forgiver has incurred definite costs associated with the initial transgression, but has only probabilistic opportunities to reap benefits from a continued relationship with the forgiven transgressor. In contrast, the feigned forgiver has made peace with the transgressor (as far as the transgressor is concerned), but remains prepared to exact revenge, if a repeat transgression occurs or if the benefits of maintaining the relationship become lower than the costs associated with exacting revenge on the transgressor.

The feigned forgiver incurs the same initial costs as the genuine forgiver associated with the original transgression, but retains multiple avenues for securing benefits from a continued relationship with the transgressor, while also remaining poised to exact revenge, perhaps pilfering resources and other benefits from the transgressor in the process. The feigned forgiver but not the genuine forgiver can receive benefits from the continued relationship with the transgressor and also can exact revenge when the transgressor least expects this, reaping the rewards of revenge while minimizing the costs incurred.

We present examples from literature and television to illustrate the differing trajectories of the "genuine forgiver" and the "feigned forgiver." In Leo Tolstoy's *Anna Karenina*, the protagonist Anna has a long-term sexual affair. Her husband, Karenin, discovers the affair but eventually forgives her (forgoing his attempts to secure a divorce) after Anna suffers a brief medical emergency. Following the reconciliation, Anna reignites the relationship with her lover and the two flee Russia. Karenin is left to raise their children alone. Karenin suffers dearly as a consequence of genuinely forgiving his transgressor. Not only does Karenin lose his long-term partner to a rival, but also he must now raise their children alone and contend with a dramatic loss of social status as a consequence of Anna's infidelity and desertion.

In the second season of *The Sopranos*, mafioso Tony Soprano discovers that his close friend, Salvatore, is a confidant for the FBI. In denial and not wanting to hurt a valuable ally, Tony maintains his relationship with Salvatore until he discovers more incriminating evidence against him. Under the pretense of looking at a new boat, Tony tricks Salvatore into lowering his guard and the two go out to sea. Tony murders Salvatore and dumps his body in the water. In this example, we see the victim maintaining a relationship with the transgressor until the transgressor is revealed to be irrevocably untrustworthy. At this point, the initial victim (Tony) lulls the transgressor (Salvatore) into a false sense of security, and then exacts revenge with little cost to himself.

In conclusion, McCullough et al. propose the existence of adaptations that motivate genuine forgiveness when the potential benefits of continuing the relationship with the transgressor outweigh the costs incurred by the transgression. The costs incurred are definite, however, whereas future benefits associated with forgiveness are only probabilistic. This situation exposes the forgiver to cheating in the form of repeat transgression. Psychological adaptations motivating genuine forgiveness are therefore unlikely to evolve. What is more likely to evolve are adaptations that motivate feigned forgiveness, lulling the transgressor into reconciliation and then exacting revenge when he least expects it.

Affective antecedents of revenge

doi:10.1017/S0140525X12000556

Kieran O'Connor[a] and Gabrielle S. Adams[b]

[a]*Graduate School of Business, Stanford University, Stanford, CA 94305;* [b]*Organisational Behaviour, London School of Business, London NW1 4SA, United Kingdom.*

koconnor@stanford.edu gadams@london.edu

http://www.london.edu/facultyandresearch/faculty/search.do?uid=gadams

Abstract: We propose that revenge responses are often influenced more by affective reactions than by deliberate decision making as McCullough et al. suggest. We review social psychological evidence suggesting that justice judgments and reactions may be determined more by emotions than by cognitions.

McCullough et al. posit that revenge is an evolved cognitive mechanism meant to reduce the likelihood that a victim will incur future harm. Specifically, they discuss the ways in which this rational system (a deliberative calculation of the Welfare Tradeoff Ratio) operates in individuals and modern cultures. However, although they include affect in their model as the "proximate motivating system" (sect. 3.1.3, para. 10), their model assumes that the role of affect in decision-making is superseded by the cognitive mechanism responsible for calculating costs and benefits (see sect. 3.1.3, para. 11). We review social psychological research suggesting that revenge is far less deliberative than the authors argue and results from more emotionally-driven reactions and biased perceptions. In particular, such research has shown that emotions are the primary drivers of decisions to seek revenge, and that individuals are insensitive to deterrence motives.

Individuals frequently respond to injustice with anger, outrage, and a heightened motivation to retaliate against the perpetrator (Darley & Pittman 2003; Carlsmith & Darley 2008). Inasmuch as deliberative cognitions play a role in the processes leading to revenge, they are likely to be influenced by the decision-maker's emotions (Schwarz 2000; Schwarz & Clore 1983).

Some theorists have argued that anger is the primary emotion involved in responses to injustice, leading to an increased tendency to support the punishment of offenders (Darley & Pittman 2003; J. S. Lerner et al. 1998). Anger not only increases tendencies toward punitive action but also reduces cognitive processing, leading to heuristic-based judgments with stronger attributions of blame and desires for retribution (Goldberg et al. 1999). Moreover, anger induced by events unrelated to a transgression can increase the desire for punishment, suggesting that even incidental emotions shape justice judgments and responses. In these instances, "hot" emotional states override deliberate cognition and decision-making (Loewenstein 1996). Avengers who are victims of the transgression grant themselves a special claim to justice (Zitek et al. 2010), and are particularly likely to experience such emotional states.

Consistent with research on the tension between affect and deliberation (Zajonc 1968), research on moral intuitions suggests that individuals express moral disapproval and outrage primarily as a result of emotional reactions, and only later develop and express cognitive justifications (Haidt 2001). More recent work has suggested that even such apparently ideological variables as political attitudes (Graham et al. 2009) and voting behavior (Inbar et al. 2012) are influenced by individuals' propensity to experience disgust. Similarly, research from outside the moral domain has established that when asked to elaborate on their decisions, individuals are quite inaccurate when it comes to describing the reasons for their actions (Nisbett & Wilson 1977). In sum, any cognitive justification that individuals claim motivates revenge (e.g., deterrence) may be post-hoc and less predictive of their decision than their emotional reactions.

Taken together, these findings run counter to the idea that revenge is based on a rational economic calculus meant to reduce the future likelihood of repeated wrongdoing. Although individuals may believe that they seek revenge for primarily utilitarian reasons, this research suggests that they may actually be relatively insensitive to probabilities of deterrence and are instead driven to punish by more emotion-based motives (Carlsmith et al. 2002). Indeed, some research has supported this contention specifically for judgments of blame and punishment. Such research has found that individuals are more likely to blame and punish offenders whose intentions evoke moral outrage than those who perpetrate the same harmful acts for other reasons (Cushman 2008); this runs contrary to the deterrence perspective, according to which people should be insensitive to motives so long as the probabilities of future wrongdoing are comparable. For instance, individuals attribute more blame and causation to a transgressor (who, e.g., injured another driver in an accident) if the reason for the outcome was attributed to an undesirable behavior (e.g., speeding home to hide a vial of cocaine) versus a desirable one (e.g., speeding home to hide an anniversary present) (Alicke 1992). In other research suggesting that punishers are not driven by cognitively based deterrence calculations, individuals confronted with incontrovertible evidence that capital punishment is an ineffective deterrent nonetheless maintained strong support for capital punishment (see Carlsmith et al. 2002; Ellsworth & Ross 1983; Vidmar & Miller 1980). Thus, individuals seem to be motivated to punish based primarily on the degree of their outrage, not the potential incapacitation of the offender (Darley et al. 2000; Kahneman et al. 1998).

Certainly, any decision-making process is likely to contain both cognitive and affective processes (Schwarz 2000). Indeed, there are contexts in which revenge may be a more deliberative "cold" process (Bies & Tripp 1996), and we do not claim that revenge is always and only emotionally driven. However, we do maintain that victims' affective reactions to wrongdoing or injustice, such as anger, outrage, and resentment, are likely to be stronger predictors of their revenge responses than are their cognitive calculations of future benefits and costs.

In light of the research reviewed above, we believe McCullough et al.'s analysis does not capture the full extent to which emotions serve as predictive causes of revenge and punishment. While the authors remind us that evolved mechanisms may not manifest themselves in identical ways for which they were selected, they do provide examples of present-day behavior (e.g., recent data from economic games) as support for their evolved mechanism. In contrast, the social psychological research on emotions, moral decision-making, and punishment which we have reviewed paints a picture of revenge as more of a hot, reflexive reaction than a cold, judicious decision.

Third parties belief in a just world and secondary victimization

doi:10.1017/S0140525X1200043X

Farzaneh Pahlavan

Laboratoire de Psychologie des Menaces sociales et environnementales, EA 4471, Institut de Psychologie, Université Paris Descartes, 92774 Boulogne-Billancourt Cedex, France.

farzaneh.pahlavan@parisdescartes.fr

Abstract: This commentary focuses on how third parties impact the course of acts of revenge based on their world views, such as belief in a just world. Assuming this belief to be true, the following questions could be asked: (a) What are the consequences of a third party's worldview in terms of secondary victimization? (b) Are bystanders actually aware of these consequences? (c) If so, then why do they let it happens?

McCullough et al. define revenge as a response to harm-imposing, *"to deter cost-impositions* […] *in the future"* (sect. 2.2.4, para. 1). As straightforward and clear-cut as this definition is, it is still reasonable to raise the question of who should evaluate the social reality of the *response*. The social reality/utility of an *event* depends on the manner and the social context in which the event is constructed by individuals. Being highly involved, both the enactor and the receiver of revenge are motivated by the same self-serving judgments, leading both of them to feel that the social order is under threat. So they have to penalize the transgressor for what he or she has done. In contrast, the perspective of third parties is hypothetically the most objective. This is particularly true if the third party is not directly or indirectly involved and is fully informed of the intervening processes that connect the instigating condition to the revengeful act.

However, judgments of third parties may also depend on their activated worldviews. If they believe in the social-psychological utility of revenge, such as its cathartic effect (psychological closure and cessation of mental rumination), then their attention might be focused on the course of revenge. This action-oriented process leads them to disengage from situational threats (Kuhl & Koole 2004). However, if they do not believe in the social/psychological utility of revenge, they are forced to deal with it in other ways, for example, by deciding that, in a just world, it is useless to worry about the negative consequences of a person's actions. After all, "people get what they deserve and deserve what they get" (Lerner 1987).

Individuals who believe in an appropriate fit between *getting* and *deserving* can therefore move on and not think further about revenge. But research often shows that revenge can increase rumination related to one's ability to deter further transgressions (Carlsmith et al. 2008). In both cases, however, people are forced to make a rational analysis of revenge, evaluating its pros and cons, to justify any punitive decisions they make and deal with the consequences.

Researchers have distinguished two types of punishment: social *retribution* and social *deterrence*, specifically when people are exposed to senseless violence (Rucker et al. 2004, p. 679). Consequently, the motive for retribution reflects deeply held beliefs about justice, and punishment is driven by a desire to see individuals pay for their deeds. The motive for deterrence is to discourage future threats, and punishment in this case is based on a profound desire to minimize the likelihood of transgression. Hence, witnessing senseless violence may lead bystanders to impose costs on the transgressor for his or her deeds or to deter future wrongdoing in order to restore justice, at least in regard to victims.

Nonetheless, research has shown that victims of aggression and revenge must cope not only with the negative consequences caused directly by the event (primary victimization; Brickman et al. 1982), but also with the negative judgments (secondary victimization) made by others (hetero-victimization) or even by themselves (self-victimization). Secondary victimization seems to be a direct consequence of the Justice Motive, first introduced in Lerner's (1980) "belief in a just world" theory (BJW) to describe people's deep-seated need to believe that the world is a fair place. To preserve this perception of the world, people are highly motivated to abide by moral norms and to protect them from disconfirming evidence. Witnessing others' unjust misfortunes may lead people to help victims in order to deter injustice, or to adopt cognitive strategies for denying the injustice (Lerner & Miller 1978). People's inalienable commitment to justice therefore constitutes one of their sacred values, causing them to react with outrage and anger whenever this value is violated (Tetlock 2003).

Holding such a belief is congruent with the idea that the victim may not be innocent. Application of the *deservingness* strategy thus sometimes leads people to blame or derogate victims. Interestingly, the absence of an opportunity to blame or disparage the victim can lead people to a *person identification* (perception of a

unit relation) or a *position identification* (perception of a non-unit relation) with the victim, the latter corresponding to a less compassionate understanding of the victim (Lerner et al. 1976). In addition, by suggesting a distinction between *other-BJW* (the world is fair to others; Sutton & Winnard 2007) and *self-BJW* (the world is fair to me; Correia & Dalbert 2008) research has shown that *other-BJW* is a better predictor of negative attitudes toward victims.

If we consider interpersonal relationships and differences/similarity as a form of social distance, then threats to the goodness of fit between *getting* and *deserving* lead to psychological distancing from victims, by moving them away from an individual's direct experience (Liberman et al. 2007). Psychological distance involves self/other differentiations, which, through *position identification* and also *other-BJW*, activate higher-order cognitive strategies such as the use of language that is abstract (reconstructing inappropriate behavior through moral justification) and decontextualized (euphemistic), fostering a deeper sense of distance from the victims (dehumanization of the victim). In addition, such cognitive reconstructions dynamically change the event's construal level without any subsequent modification of the individual's moral standards. For Bandura (1999), these psychological mechanisms prepare people to be dogmatic or punitive, and to quickly start looking for excuses, such as anger and outrage, for having abandoned their bedrock commitment to justice. When people blame others, they are not inclined to evaluate the correctness or accuracy per se. They care only about justifiability, a profoundly relational construct that depends on the identity of the protagonists and their evaluative standards, regardless of whether it captures essential aspects of reality.

Thus, although people may be unaware of the intra-individual effects or functions of the BJW, there seems to be no doubt about its social functions in terms of justification of interpersonal relationships. This supplies a good reason for switching off one's conscience, which equates to moral disengagement.

Adaptationism and intuitions about modern criminal justice

doi:10.1017/S0140525X12000568

Michael Bang Petersen

Department of Political Science and Government, Interacting Minds Centre, Aarhus University, DK-8000 Aarhus C, Denmark.

michael@ps.au.dk

http://pure.au.dk/portal/en/michael@ps.au.dk

Abstract: Research indicates that individuals have incoherent intuitions about particular features of the criminal justice system. This could be seen as an argument against the existence of adapted computational systems for counter-exploitation. Here, I outline how the model developed by McCullough et al. readily predicts the production of conflicting intuitions in the context of modern criminal justice issues.

An adaptationist framework predicts that people's intuitions regarding modern criminal justice stem from psychological mechanisms designed for countering the evolutionarily recurrent equivalent, exploitation (Petersen et al. 2010). By implication, the model of evolved counter-exploitation mechanisms described by McCullough et al. should predict core features of criminal justice intuitions.

Most of the previous applications of the adaptionist framework to criminal justice focus on people's perceptions of crime seriousness, arguing that they emerge from evolved intuitions about harm (e.g., Robinson et al. 2007). In making this argument, researchers emphasize the high levels of consensus among individuals and countries on the relative seriousness of different crimes. If widespread agreement is a necessary product of the operations of evolved mechanisms, we would conclude that such mechanisms play little role in the production of the intuitions most relevant to McCullough et al.'s model: that is, intuitions about whether punitive or reparative goals should be given priority in criminal justice. Hence, in criminological research, consensus that people have incoherent or "mushy" criminal justice intuitions has emerged from observations of profound disagreement between individuals, and in different situations about the prioritization of punitive and reparative sanctioning goals (Cullen et al. 2000; Roberts & Hough 2002; Roberts & Stalans 2004).

McCullough et al.'s model emphasizes (1) the existence of computational systems designed to implement revengeful as well as forgiving strategies, and (2) computations of the social value of the exploiter as the key regulator of the activation of these strategies (see also Petersen et al. 2010). These two features, I argue, render disagreement between individuals and in different situations an unavoidable outcome of the operation of our evolved counter-exploitation intuitions.

This relates to the fact that the social value of others is both self-specific and target-specific. One target is the self's kin, another is not; in one case, the target and the self share a history of cooperative interactions, in another case they don't; one target is part of the same group as the self, another is not. And so on. Accordingly, different selves will compute the social value of the same target differently, and a single self will compute the social value of different targets differently. In the face of a specific exploiter, some individuals will therefore experience punitive sentiments, while others will experience reparative sentiments.

In modern mass societies, an individual's personal welfare will rarely be affected by how the state sanctions a single criminal, regardless of that criminal's social value. Nevertheless, given the evolutionary importance hereof, our minds should be designed to automatically pick up on ecologically valid cues to the presence of social value (e.g., expressions of remorse, shared ethnic background, lack of criminal record) and activate punitive and reparative sentiments accordingly. In modern criminal justice, different sanctioning types are aligned with these different sentiments to varying degrees. Cost-imposing sentences such as prison are aligned with punitive sentiments, whereas rehabilitation schemes are more aligned with reparative sentiments (e.g., Cullen & Gendreau 2000). Accordingly, different individuals will prioritize different sanctioning types in the face of the same criminal (e.g., depending on the match between the respective ethnic or racial backgrounds of the self and the criminal), and the same individual's priorities will change in the face of different criminals (depending on the specific cues surrounding each case).

This simple observation readily integrates the adaptationist framework and the demonstrations of disagreement within individuals towards different crimes and between individuals towards the same crime. According to this interpretation, lay intuitions are not "mushy," and disagreement is driven by the existence rather than lack of principled intuitions. If valid, we should find that whether individuals agree or disagree about appropriate sanctions is predictable from whether there is agreement or disagreement with respect to the perceived future social value of the criminal. Some evidence for this link already exists (Burnette et al. 2012; Lieberman & Linke 2007; Petersen et al. 2012).

Modern criminal justice intuitions have been deemed "mushy" for more than just these kinds of disagreement. In particular contexts, research has also demonstrated how people find it highly difficult to prioritize either reparative or punitive sanctioning goals. Instead, people concurrently express firm support for both (Cullen et al. 2000; McCorkle 1993; Roberts 1992). Such genuine ambivalence might initially appear more difficult to reconcile with McCullough et al.'s model. Yet, in order to understand the output of any computational system, we must simultaneously analyze the system's informational needs and the information offered by the environment (Gigerenzer et al. 1999). As McCullough et al. emphasize throughout their article,

modern environments do not necessarily deliver the cues that the revenge/forgiveness systems require to operate adaptively. The context within which modern criminal justice intuitions emerge offers a case in point. We evolved to function within small-scale groups and react towards specific exploiters. Modern political debates concerning criminal justice, in contrast, proceed within anonymous mass societies and are general rather than specific in the sense that they are often about how to react towards *all* criminals. In this context, our counter-exploitation psychology would be activated by the features exhibited by crime, but the lack of ecologically valid cues about the social value of the relevant criminals would make it difficult for this psychology to execute properly and, in particular, to up-regulate either punitive or reparative motivations at the expense of the other. Some evidence for this proposition already exists, as people's difficulties in prioritizing between punitive and reparative goals seem to be restricted to general information-sparse decision-making contexts: When specific cues are directly available, unequivocal criminal justice intuitions emerge (Petersen 2009).

Although the previous applications of the adaptionist framework have emphasized the shared nature of criminal justice intuitions, the widespread existence of conflicting intuitions about the prioritization between punitive and reparative goals should not be taken as evidence against the existence of adapted mechanisms for generating such intuitions. Hence, the model developed by McCullough et al. readily predicts the production of conflicting, changing, and ambivalent intuitions about this particular feature of modern criminal justice.

The elementary dynamics of intergroup conflict and revenge

doi:10.1017/S0140525X1200057X

David Pietraszewski

Department of Psychology, Yale University, New Haven, CT 06520-8205.
david.pietraszewski@yale.edu

Abstract: The psychology underlying revenge in an intergroup context is built around a small handful of recurrent interaction types. Analyzing the cost/benefit calculations of each agent's role within these interaction types provides a more precise way to characterize intergroup conflict and revenge. This in turn allows for more precise models of the psychology of intergroup conflict to be proposed and tested.

In their target article, McCullough et al. re-conceptualize the familiar concepts of revenge and forgiveness in terms of classes of evolutionarily recurrent situations involving potential costs and benefits, around which proximate psychological mechanisms are designed. Here, "costs" and "benefits" are not synonyms for preferences or desires, but instead correspond to classes of outcomes which relate to differential reproductive success over multiple generations. This approach sheds new light on old constructs, reframes the questions being asked, and provokes new and clear directions for future research. While the target article focuses on revenge and forgiveness, this same perspective also suggests new ways to think about intergroup conflict. For example, intergroup conflict and revenge can be understood as a sequence of triadic interactions, of which there are only four types. Following Strayer and Noel (1986), these are as shown in Figure 1:

In an Alliance, two agents attack or impose costs on a third agent: A attacks B, and C also attacks B. In Defense, one agent attacks a second, and a third agent responds by attacking the aggressor: A attacks B, C then attacks A. In Generalization, one agent attacks two others: A attacks B and then also C. Finally, in Displacement, one agent attacks another, and that attacked agent responds by subsequently attacking a third agent: A attacks B, B then attacks C. Any instance of intergroup conflict or revenge will involve a particular concatenated sequence of these interactions. What determines this sequence will be the result of cost/benefit decisions on the part of each actor.

This taxonomy of triadic conflict has served descriptive purposes for decades. But from an adaptationist perspective, this taxonomy can also be understood as describing recurrent classes of situations that humans have encountered. Moreover, for each interaction type, the perspective of each agent can be analyzed.

For example, consider (1) *Alliance*: A should calculate the likelihood that C will ally with him in determining whether or not to initially attack B. A should also consider the consequences of C also attacking B. C needs to calculate the value of allying with A against B. C may do this to ingratiate himself to A, or owe something to A, or be differentially allied with A. Or, C may have a poor relationship with B, and take advantage of this opportunity to impose a cost on B. B should represent the cost imposed by both agents, and should also consider the pairwise relationship comparisons: B with A, B with C, A with C in determining what to do next.

Or, consider, (4) *Displacement*: B may act because A has an interest in C, and by imposing a cost on C, B is intending to indirectly impose a cost on A. Or, even if A has no interest in C, B's action towards C may cause C to appeal to A to take B's welfare into account (because A's aggression towards B is now yoked to B's aggression towards C, such that if A aggresses again, C will be hurt again). The potential cost to B of doing this is that it may cause A and C to unite against B. This is less of a problem if it is already likely that C would have come to A's defense. A will represent that B has imposed cost on C, and will

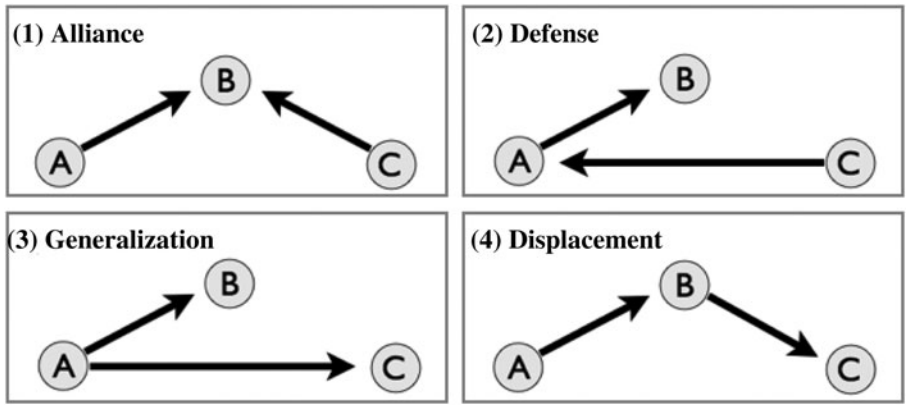

Figure 1 (Pietraszewski). The four types of triadic conflict. Arrows denote attack/cost imposition. All instances of intergroup conflict and revenge are built up out of these interaction types. (Adapted from Strayer & Noel 1986).

subsequently calculate the effect on C. To the degree A has an interest in C, or that C will retaliate towards A, A should consider B's action a cost. C will represent that B imposed a cost.

The proximate psychology of multi-person conflict, revenge, and retaliation is built around these recurrent interaction types. Considering each agent's perspective within these triadic interactions therefore allows one to reframe the vague and difficult question, *What is the proximate psychology of intergroup conflict?*, into sets of deductively richer and tractable questions, such as: *What are the classes of situations in which an agent finds themselves in role B during an alliance event? What cost/benefit considerations would that agent need to consider?* And, *What are the on-the-ground cues that would facilitate identifying these situations?*

From this perspective, the folk construct "group" can be understood as classes of relationships between agents which cause them to be more likely to be in particular roles within these interaction types. What it means to be members of a group, as far as the design of the proximate psychology may be concerned, is to be in roles A & C in Alliance, B & C in Defense and Generalization, and A & C in Displacement. These roles would be part of the cue structure on the input end (i.e., observing these roles allows one to deduce an intergroup conflict is unfolding, and also who is allied with whom), and also be embodied in the motivational and representational changes on the output end (being allied or in a group with someone makes it more likely one will execute the behaviors that correspond to these roles). This would be true of the psychologies driving the decisions of the actors within the group context, as well as the psychologies of third parties who are forming expectations and updating their representations of the groups and their members.

This analysis suggests a tentative answer to a question posed by McCullough et al., of whether the psychology that governs the operation of revenge systems also evolved to regulate behavior in intergroup contexts. In many respects, the proximate psychology governing intergroup conflicts is probably interestingly different from systems primarily designed around dyadic revenge (different because it requires triadic, rather than dyadic calculations, and because its cue structure and subsequent behavioral and motivational responses will likely be somewhat different). However, even if the proximate psychologies are interestingly different, they are both examples of phenotypic design to contingently respond to direct or indirect costs imposed by other agents.

Revenge: An adaptive system for maximizing fitness, or a proximate calculation arising from personality and social-psychological processes?

doi:10.1017/S0140525X12000441

Michael Potegal

Division of Pediatric Clinical Neuroscience, Department of Pediatrics, University of Minnesota Medical School, Minneapolis, MN 55455.

poteg001@umn.edu

Abstract: Revenge appears among a "suite" of social interactions that includes competition, alliance building (a prerequisite for tribal revenge raids), and so forth. Rather than a modular "system" directly reflecting evolutionary fitness constraints, revenge may be (another) social cost-benefit calculation involving potential or actual aggression and proximately controlled by individual personality characteristics and beliefs that can work against fitness.

In focusing upon revenge, McCullough et al. have raised an interesting topic with broad cultural and social significance. If anything, their article may underestimate its importance in, for example, the contemporary workplace (e.g., Tripp & Bies 2010). One unnoted aspect of revenge is its powerful and seemingly universal role in motivating warfare in preliterate cultures in which homicide rates can rise well beyond those in industrialized societies. In the recent past, the Waorani of the Ecuadorian rainforest (Beckerman et al. 2009) and the Gebusii of lowland New Guinea (Knauft 1987) were each locked into repeating cycles of within-group revenge murders; both may have been on their way to killing themselves off before being rescued by outside intervention.

The authors clearly and commendably distinguish between the ultimate shaping of behavior by its effects on fitness and its proximate behavioral/motivational instantiations. They argue that revenge is an identifiable (presumably modular) system whose characteristics directly reflect the fitness pressures that shaped it. While I agree that revenge is important, I suggest that (1) it is just part of a suite of proximate behavioral/motivational mechanisms that calculate social cost–benefits involving potential or actual aggression, and that (2) these powerful mechanisms shape revenge in ways that are independent of, or even antithetical to, whatever fitness pressures might have given rise to them. Thus, the phenomenon of people reciprocally locking themselves into lethal revenge cycles is a major and fairly common "misfire" (the authors' term) of proximate mechanisms in the face of supposed ultimate maximization of fitness. Four other caveats and a comment on tantrums follow.

The reinforcing value of aggression. In a drive-by allusion, McCullough et al. acknowledge, but do not give due weight to, the intrinsically reinforcing aspects of aggression as a primary reward in exacting revenge. Several sorts of evidence indicate that some men enjoy aggression. Historic episodes, when subcultures enjoyed "recreational" fighting, include 15th century Venetian bridge fights, 19th century Irish "faction fights" ("Donnybrook" refers to a seasonal Irish fair featuring big fights), and end-of-season brawls among American loggers (Ingle 2004). Individuals in subcultures in which fighting is a norm, for example, some English soccer fans (Buford 1992) and pub-goers (Graham & Wells 2003), continue to provide subjective reports of pleasure in fighting; but even in samples not selected for fighting, a modicum of pleasure in aggression is reported (Ramírez et al. 2005; cf. Nell 2006).

Of course, beyond any intrinsic reward in activating the neural circuitry of aggression, its reinforcing value may be shaped by, for example, pride and self-esteem in fighting and winning and/or relief of tension and fear of harm. Developmentally, however, boys' pleasure in aggression appears as a major sex difference at least as early as age four (Benenson et al. 2008). Phylogenetically, aggression's reinforcing value has been demonstrated across many species; fish swim through rings, birds peck at keys, and mice and/or rats poke their noses in holes, press bars, and run across electrified grids just to attack a conspecific (e.g., May & Kennedy 2009).

Anger versus rumination. If anger were its driving affect, revenge would be short-lived. Surveys suggest that ordinary episodes of anger persist for no more than about 30 minutes, usually less (Potegal 2010). Unless revenge is taken in the moment, one must invoke anger transformed into rumination, which is a highly elaborated cognitive activity that includes detailed (sometimes fanciful) planning for the future. This accords with the advice of boxing coaches, military generals, and others who urge on their fighters the proverbial "Revenge is a dish best eaten cold." At a pathological extreme, rumination on revenge that eventually results in a "catathymic crisis" can disrupt the would-be avenger's life for weeks, months, or even years.

Personality. In the context of anger, the frequency and intensity of revenge is likely closely associated with an individual's general level of hostility as indicated by, for example, standard measures of hostility or Big 5 (dis)-agreeableness, as the authors and others have shown. If so, how much variance in individual behavior is explained by cost-benefit analysis of particular situations versus

general personality characteristics? Although individual hostility, enjoyment of aggression, risk-taking, and so forth could be treated as parameters in equations predicting revenge, if there are, for example, many individuals who never take revenge and/or many who always do, no matter how slight the offense, then revenge depends more on proximate personality than on ultimate causation as reflected in situational demand.

Witchcraft murders – Revenge against kin. Witchcraft trials in Europe and North America ended more than 250 years ago, but in parts of India, Sub-Saharan Africa, and elsewhere, thousands of people have been exiled or killed for being witches in the last three decades (Federici 2010). A witchcraft believer who experiences misfortune or illness will suspect that the culprit is someone who harbors ill feelings toward him. In small communities, these may well be family members (kin or affines) because quarrels with them are the most likely. Remarkably, many children have been abandoned, injured, or killed by people including their parents (Adinkrah 2011; UNICEF Report 2010). Witchcraft accusations among kin, especially against children, contradicts biological fitness expectations and is more consistent with proximate psychological processes and beliefs.

Finally, the parenthetical comment that tantrums are among behaviors that children impose on parents "with impunity" requires more attention to detail. Across quite different groups of children, tantrums are composed of one set of behaviors reflecting anger (e.g., shouting, hitting) and a second set reflecting distress-sadness (e.g., crying, whining, and comfort seeking, Green et al. 2011; Potegal & Davidson 2003; Potegal et al. 2003; 2009). This differentiation has been replicated (Giesbrecht et al. 2010). Anger and sadness/distress have different temporal profiles. Angry behaviors peak early in tantrums, then fade while sadness/distress behaviors remain relatively constant throughout. Thus, tantrums end with children's sadness/distress and comfort seeking. Functionally, the child's terminal sadness pulls for parent comfort-giving, that is, behavioral "forgiveness." Thus, child sadness and corresponding parent comfort-giving help end the tantrum and repair social bonds that were just strained by the child's anger (Potegal 2000; Potegal & Davidson 1997).

Applying the revenge system to the criminal justice system and jury decision-making

doi:10.1017/S0140525X12000581

S. Craig Roberts[a] and Jennifer Murray[b]

[a]*Department of Psychology, School of Natural Sciences, University of Stirling, Stirling, FK9 4LA, Scotland, United Kingdom;* [b]*Nursing, Midwifery, and Allied Health Professions Research Unit, University of Stirling, Stirling, FK9 4LA Scotland, United Kingdom.*

craig.roberts@stir.ac.uk

jennifer.murray@stir.ac.uk

http://www.psychology.stir.ac.uk/staff/staff-profiles/academic-staff/craig-roberts

http://stir.academia.edu/JenniferMurray/About

Abstract: McCullough et al. propose an evolved cognitive revenge system which imposes retaliatory costs on aggressors. They distinguish between this and other forms of punishment (e.g., those administered by judges) which are not underpinned by a specifically designed evolutionary mechanism. Here we outline mechanisms and circumstances through which the revenge system might nonetheless infiltrate decision-making within the criminal justice system.

Applying the revenge system to the criminal justice system and jury decision-making. The proposed evolved cognitive revenge system serves two purposes: to discourage an aggressor from imposing future costs to their victim, and to encourage the aggressor not to withhold future benefits from the victim. McCullough et al. suggest that the revenge system is similar to the criminal justice system (CJS) in some respects, but is fundamentally distinct from institutionally organised forms of punishment such as the ones administered by judges. However, we believe that, under certain circumstances, the revenge system may not be as distinct from the CJS as McCullough et al. suggest. In particular, it may be relevant to the applied context of jury decision-making, where a group of lay people are tasked with deciding whether or not a person is guilty of committing a crime (this can be wholly, partially, or not guilty). In light of growing recognition of the potential for applying evolutionary insights to specific issues in modern society (e.g., Roberts 2012), we here discuss the ways in which the proposed revenge system can be applied to the CJS.

The main distinction between the evolved revenge system and institutionally administered punishment lies in their respective foci. While the former is characterised by a mechanism designed to deter cost-impositions or benefit-withholdings in the future (as McCullough et al. argue in the target article), and is underpinned by emotional motivation (e.g., Lapsley 1998), the latter is based on socially developed constructs of justice (Ho et al. 2002; Price 1997). The CJS therefore aims to allocate suitable and appropriate punishment in an emotionally detached fashion, according to pre-identified guidelines and societal norms (Lerner 2003).

A further distinction between the revenge system and jury decision-making lies in the level of personal involvement. Whereas the revenge system is discussed in the context of a victim-aggressor relationship, where the costs and benefits are highly personal and relevant to the individuals directly involved, CJS decision-making involves almost no true personal involvement, as no previous (or probable future) relationship normally exists between the aggressor and legal representatives or jurors.

There are, however, occasions where the revenge system may infiltrate the CJS. This is because punitive decisions in the CJS cannot always be wholly extricated from emotional influence (Ho et al. 2002; Murray et al. 2011). Individuals who hear intimate details about a case may involuntarily become emotionally involved, especially where there has been a high cost to the victim (e.g., extreme violence, sexual assault), leading to empathy for the victim (Tsoudis 2002). Jurors, in particular, may be influenced by the emotional re-telling of an incident, as they are relatively unlikely to have experiences of such cases in their day-to-day lives and have little to no formal training in legal processes. Through increased emotional involvement and empathy, an "emotionally involved" juror may come to view the costs of the crime in a personal manner, seeing the benefits of punishment in a similar way to the victim, leading to the desire for "vengeance" and stronger punitive sentiment (Ho et al. 2002; Lapsley 1998; Murray et al. 2011). Thus, because of the evolutionary link between emotion and vengeance, the proposed revenge system may well be applicable to decision-making by jurors (and possibly others), at least in cases where emotional valence and cost to the victim is high.

Understanding the likely circumstances under which the revenge system may be activated within institutionally administered punitive decision-making is a necessary step towards making such processes more balanced and fair. Activation of the revenge system is less likely in cases where the cost to the victim, and therefore levels of empathy, are relatively low (e.g., petty crimes, which constitute the majority of cases). In such instances, punitive decisions may be better explained and guided by considering socially defined justice processes and norms. In contrast, as we have discussed, the revenge system cognitive architecture may be suitable for understanding decision-making in more emotionally valenced cases and especially when jurors are involved.

Furthermore, research is urgently needed to understand potential between-individual differences in susceptibility to emotional involvement and its corollary effects on judgements and punitive decision-making. Our recent unpublished data, for example, suggest that personality traits predict levels of anger at transgressors

and the desire to punish them. Another example is potential gender differences in emotional involvement. In a scenario involving transgression in a public-goods game, men expressed a greater desire to punish "cheats" than did women (O'Gorman et al. 2005). Men also showed different empathy-related activation responses than women in response to individuals who played unfairly in a study of another economic game, and were more likely to express desire for revenge and to favour physical punishment (Singer et al. 2006). Furthermore, it is now well-known that evolutionarily-relevant characteristics of the defendants (such as their sex and attractiveness), and shared characteristics between defendants and jurors (such as race or sexuality, triggering in-group/out-group prejudice), influence punitive sentiment and leniency or harshness in sentencing (e.g., Abwender & Hough 2001). These findings may be explained through involuntary activation of the revenge system's cognitive architecture.

Although there are clear distinctions between the proposed revenge system in its current form and institutionally administered punishment decisions, the revenge system may be useful in explaining punitive decision-making in a number of applied contexts in the CJS, notably where jurors are involved and emotional valence in a case is high. Through considering justice as a mediating factor alongside the already existing components of the revenge system, the theory may also be applicable at a more "socially driven," justice-based decision-making level. A final potential use for the revenge system is in future investigations of punitive decision-making in the CJS relating to between-individual differences, such as personality and gender differences, as we have discussed. Through better understanding the ways in which individuals come to their decisions about punishment, improvements to punitive decision-making processes within the CJS will be made possible.

Forgiveness is institutionally mediated, not an isolable modular output

doi:10.1017/S0140525X12000593

Don Ross

School of Economics, University of Cape Town, Rondebosch 7701, Cape Town, South Africa.

don.ross@uct.ac.za

http://uct.academia.edu/DonRoss

Abstract: McCullough et al. recognize that revenge and forgiveness jointly constitute a functional strategic complex. However, they model the halves of the complex as outputs of modules selected for regulating dyadic relationships. This is backwards. Forgiveness is a culturally evolved institution that can be exapted for use in dyadic contexts; it would be cheap talk among dyads were it not for the shadow of society.

McCullough et al. perform several useful services in their target article. They remind us that the familiar phenomenon of revenge-seeking is not a pathology for which forgiveness is a cure – so that the "progress of civilization" might be imagined as leading toward a world where forgiveness abounds and revenge goes the way of foot-binding and cigarette smoking. Forgiveness, McCullough et al. correctly stress, is part of a functional complex *with* revenge, and the complex as a whole is almost certainly maintained by selection dynamics. Seeking revenge is often welfare-promoting for an agent because it changes others' incentives toward the self (see sect. 2.2.1), ideally inducing an increase in cooperation and/or a decrease in exploitation. Most importantly in my view, McCullough et al. recognize that forgiveness is important, interesting, and indeed theoretically surprising because it seems on the surface to be cheap talk. Once I exact my revenge against you, in equilibrium you should recognize that, if our continued relationship is of any value to me, I have no incentive to continue to follow a vengeful course if you avoid the action that triggered it. And since announcing the words "I forgive you" seems costless to me, what information value could the announcement possibly add to your appreciation of our equilibrium conditions? And yet most people regard the timing and circumstances of forgiveness as matters of solemn significance.

In other writings (Ross 2004; 2005; 2006; 2007; 2012) I have suggested that a key condition for the existence of equilibrium strategy vectors that include revenge and forgiveness is the human susceptibility to shame – that is, to suffering from highly aversive emotions when one detects that information about one's norm violations may be spreading through social networks by gossip. Gossiping about a transgression is of course not the only, nor always the most efficacious, means by which revenge is carried out. However, it is arguably the most common. And, as I have argued, it is the crucial mechanism that promotes the ubiquitous existence of institutions for legitimizing forgiveness. Members of a species who depend on specialization and exchange of goods and services benefit from mechanisms that allow reputationally damaged members of economic networks to be restored to productive membership once corrective punishment is thought to have triggered policy reform on their part. Institutions for forgiveness allow the social costs of the most common acts of revenge to be massively reduced, at least when revenge is taken through harming a transgressor's reputation as opposed, for example, through harming his kneecaps. This in turn solves the much-discussed problem that arises when punishment of norm violation is costly to punishers, that agents are incentivized to free ride on the public good of norm enforcement (Guala 2012).

It is of crucial importance, on this account, that forgiveness is *institutionalized* and *public*, and operates mainly in games that involve more than two players. Of course, once institutions for forgiveness exist, and their force is internalized through socialization and enculturation of young people, they can be exapted for use in dyadic interactions. But the shadow of society is always present in such interactions. Use of a phrase such as "I forgive you" is regulated by pragmatic (Gricean) conventions that include in their implicatures the act of promising not to spread blame for the transgression at issue. Forgiveness is a species of promising, and promising is a conventional institution regulated by socially enforced norms.

In light of these considerations, it is doubtful methodology to try to develop a model of the evolutionary function of revenge and forgiveness, as McCullough et al. do, by beginning with the case of the isolated dyad. The authors go to some lengths to analytically distinguish forgiveness – in a technical sense they deliberately construct – from more complicated institutionally governed cognitive/behavioral relatives. They are compelled to go to this trouble because they want to promote the hypothesis that there are specialized evolved cognitive *modules* for revenge and forgiveness that societies and institutions can then exploit – and sometimes suppress – as networks of interactions become more complex. Such modules would support special senses of revenge and forgiveness that can be isolated from social dynamics.

I do not see that in their target article McCullough et al. produce *any* evidence for this conjecture. It is plausible that institutional and culturally evolved and stabilized revenge/forgiveness complexes depend on biologically selected dispositions to be emotionally sensitive to perceptions of signals of possible changes in one's reputation among conspecifics. Emotions related to social rejection and reconciliation are probably evolved dispositions present in all normal primates, cetaceans, canids, corvids, et cetera. Such dispositions are likely necessary aspects of the evolution of cognitively mediated sociality in general. However, once social interdependence has evolved and thrown up free rider problems, strategies based on reciprocity are solutions that any cultural dynamics are likely to find and stabilize. I see no evidence that human institutions for forgiveness couldn't be supported by general cognitive processes and should be thought to require dedicated modules. Of course, dogs and

elephants (etc.) lack resources for building forgiveness institutions. But what in their behavior should lead us to think that their cognitive processing goes beyond general dispositions to reconcile with group members following non-lethal altercations?

The hypothesis I favor is that the individual psychology of interaction is parasitic, both developmentally and theoretically, on its social psychology, economics, and political anthropology. The science should start at the aggregate level and treat individual and dyadic expressions as, respectively, exaptations (when characterizing development) and abstractions (when building game-theoretic models). For example, *avoidance* should be modeled as the special dyadic case of the primary phenomenon of ostracism from the group. McCullough et al. deliberately pursue the opposite modeling approach.

Ethics preclude testing between these hypotheses by isolating some human children from opportunities to learn forgiveness conventions. But if there are dedicated modules for computing revenge and forgiveness, we should find people suffering from neural deficits that knock out these capacities while leaving other emotionally motivated and regulated cognition intact. I am aware of no clear reports of such cases in any literature.

Revenge can be more fully understood by making distinctions between anger and hatred

doi:10.1017/S0140525X1200060X

Aaron N. Sell

School of Criminology and Criminal Justice, Mt. Gravatt Campus, Griffith University, Mount Gravatt, QLD 4122, Australia.

sell@psych.ucsb.edu

Abstract: McCullough et al. present a compelling case that anger-based revenge is designed to disincentivize the target from imposing costs on the vengeful individual. Here I present a contrast between revenge motivated by anger (as discussed in the target article) and revenge motivated by hatred, which remains largely unexplored in the literature.

The idea that hatred is an evolved adaptation has not been thoroughly elaborated (though see Waller 2004; Petersen et al. 2010; Sell 2011). I give a brief outline here of how a selection pressure distinct from those considered in the target article could design a system – hatred – that enacts revenge in response to different triggers, moderated by different variables and designed for different purposes.

Consider the following selection pressure: *The existence and well-being of another individual has some impact on your lifetime fitness; sometimes that impact is negative.* Such an individual could have a low Welfare Tradeoff Ratio (WTR) towards you (e.g., a bully); could be someone who values you highly but nevertheless imposes large costs (e.g., a flirtatious student whose inappropriate behavior threatens your marriage); or even be someone who doesn't know you exist (e.g., the person who holds the job you want). This selection pressure would design a mechanism that identifies these individuals and then deploys – when cost-effective – behavioral strategies that reduce the target's ability to impose costs by limiting interactions with the target, reducing their power, or killing them. Let us call the mechanism "hatred," recognizing that in layman's terms "anger" and "hatred" are often conflated.

According to the recalibrational theory, and consistent with the arguments made by McCullough et al. in the target article, the function of anger-based revenge is to raise the target's WTR (Sell et al. 2009; Sell 2011). The function of hatred, in contrast, is to reduce costs that emanate from another individual by isolating, weakening, or killing them. These are different designs and should generate distinct predictions about their antecedent conditions and behavioral strategies.

The triggers of hatred are theoretically distinct from those of anger. As discussed in the target article, anger is triggered by indicators that another does not value one's welfare highly. Hatred is triggered by indicators that another's existence and well-being will cause harm. Both anger and hatred can be triggered by the intentional imposition of large costs for trivially small benefits. But hatred can also be triggered by rivals for mates or status, even when these rivals hold the hateful person in high regard (i.e., have high WTRs towards the hateful individual). Targets can also be hated for the imposition of frequent small "justified" costs that do not indicate a low WTR but nonetheless bear on the hateful individual's future welfare. Such an explanation could explain the perversely high prevalence of physical aggression against elders who require living assistance (Lachs & Pillemer 2004; Pillemer & Finkelhor 1988) and disabled children who require additional investment (Westcott & Jones 1999).

Also unlike anger, hatred can motivate seemingly vengeful behavior when the target has done nothing other than *be* harmed by the hateful person. Harming someone will cause them to lower their WTR, retaliate, or spread word of the misdeed to others. This makes the original victim into a person whose existence predicts future costs for the offender. As paradoxical as it seems, subjects who were made to insult or harm others did indeed dislike them for it (Schopler & Compere 1971), but they cannot be said to have been "angry" at their victims.

The behavioral consequences of hatred and anger partly overlap, specifically with the enactment of revenge. Both anger and hatred can fulfill their functions by imposing costs on the target individual. However, if the selection pressures responsible for each emotion are distinct, then we would predict certain differences in their behavioral strategies. One difference is that hatred should be largely indifferent to apologies or signals of recalibration. In fact, apologies – to the extent that they indicate the hated person is suffering – could be experienced pleasantly and incentivize more cost infliction. This is because the "off switch" to anger and hatred are different. Anger has served its function when the target recalibrates, but hatred has fulfilled its function only when the target has been significantly de-powered, killed, or ostracized. Indeed, anger-based aggression is frequently negotiative in design with clearly demarcated starting and ending points, turn taking, escalation starting with low-cost assessments of formidability, and an understanding of "fairness" that tracks the accuracy of assessments; for example, hitting someone while they are asleep does not demonstrate your formidability and bargaining power (see Sell 2011). Revenge stemming from hatred is predicted to have none of these features.

As indicated in the target article, bargaining power (such as physical strength in men) is a predictor of anger-based aggression (Sell et al. 2009), because those with better bargaining power will have more success deploying that tactic. For anger to fulfill its function, one must confront the target and convince them that one's interests are worthy of being weighted more highly. Hatred, on the other hand, can fulfill its function without the target ever knowing the mental state of the hateful person, and by taking advantage of temporary fluctuations in bargaining power. For this reason, when faced with an individual who holds a low WTR, a person with high bargaining power can recalibrate the target (and evidence shows they do), whereas a person with low bargaining power will instead hate the target and look for subtle opportunities to impose costs or temporary shifts in bargaining power that can be used to weaken or kill the target (e.g., backstabbing, gossip, sabotage). Consistent with this, there is evidence that physical strength in men positively correlates with anger and aggression, but does not predict their tendency to ruminate or seethe over affronts (Sell et al. 2009).

The relationship between anger and hatred is complex. Anger can trigger hatred if a person resists recalibration and becomes the perpetual cost inflictor that is strategically better addressed by hatred. Conversely, a hated individual who raises his WTR may become less hated because of the benefits he will bestow.

Finally, both emotions can run in parallel, attempting to recalibrate the target's WTR while limiting their ability to impose costs. Regardless of these complications, anger and hatred appear designed by separate selection pressures with different triggering conditions, moderating variables, and behavioral strategies. Revenge reflects the operation of both adaptations and will have to be understood that way.

Revenge and forgiveness in the New South Africa

doi:10.1017/S0140525X12000453

Dan Joseph Stein,[a] Jack van Honk,[a] and George Ellis[b]

[a]*Department of Psychiatry, University of Cape Town, Groote Schuur Hospital, Observatory 7925, Cape Town, South Africa;* [b]*Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7701, Cape Town, South Africa.*

dan.stein@uct.ac.za **jackvanh@gmail.com**
george.ellis@uct.ac.za

Abstract: Insofar as South Africa underwent a rapid transformation from apartheid to democracy, it may provide a unique laboratory for investigating aspects of revenge and forgiveness. Here we suggest that observations and data from South Africa are partially consistent with the hypotheses generated by MCullough and colleagues. At the same time, the rich range of revenge and forgiveness phenomena in real-life settings is likely to require explanatory concepts other than specialized modules and their computational outputs.

The hypotheses generated in McCullough et al.'s article can be tested in multiple ways and contexts. Having worked in the interesting context of the new South Africa (Allan et al. 2006; Kaminer et al. 2001; Stein et al. 2008), we read the target article from the vantage of whether observations and data about revenge and forgiveness in this setting might be relevant. Insofar as South Africa underwent a rapid transformation from an apartheid system characterized by racial discrimination to a democratic dispensation characterized by universal human rights, it may provide a unique laboratory for investigating certain aspects of the psychology of revenge and forgiveness. We note McCullough et al.'s warning that their model is not intended to apply to groups, but groups are of course comprised of individuals.

A first point to note is that in South Africa, with the advent of the new democratic dispensation in 1994, a decision was made that instead of retributive justice (i.e., punishment), there would be reparative justice (as embodied by a "Truth and Reconciliation Commission") (Asmal et al. 1996; Stein 1998). One key rationale was that it simply wasn't possible to ensure retributive justice, given that the institutions responsible for implementing justice after 1994 had not yet democratized. It is notable that McCullough et al. posit from basic evolutionary principles that punishment is less effective when the costs of punishment are high. This was certainly the case in post-democratic South Africa.

Nevertheless, there were many who objected to the Truth and Reconciliation Commission (TRC) over precisely this issue. In particular, families of those who had been victims of gross human rights violations were vehemently opposed to the idea that perpetrators would not be formally prosecuted and punished. Notably, McCullough et al. provide a model which predicts that revenge is more likely when a kin group member is involved. Thus, revenge may be well be sweeter for some than for others (de Quervain et al. 2004; Stein & Kaminer 2006). Although the courts are not always seen as active in the South African setting, in the particular instance of the TRC the judicial system was clearly active (albeit administering reparative rather than retributive justice), predicting, per McCullough et al., less need for retaliatory feelings in the majority of observers (Stein et al. 2008).

Indeed, a second point is that forgiveness levels appear to be moderately high in all sectors of the South African population in the aftermath of the TRC (Stein et al. 2008). This may well be consistent with McCullough et al.'s model, which predicts that shared interests, similar values, and many opportunities for mutually beneficial interactions are good candidates for forgiveness. Certainly, despite clear racial heterogeneity in South Africa, there are many values that are held in common by the population, including religious values, and despite considerable geographical separation of races, there are many opportunities for interaction (e.g. in domestic and commercial settings). Relationships are in many ways "valuable," and therefore considerable conciliatory behaviour could perhaps have been predicted.

Nevertheless, there have been many in South Africa, including participants in the TRC, with low levels of forgiveness for past violations (Stein et al. 2008). Again, there are a number of potential explanatory variables. McCullought et al. note that apology is important, and the South African data confirm this (Allan et al. 2006). McCullough et al. cite meta-analytic findings that women score higher on self-reports of tendencies to forgive, but the South African data suggest that men are more likely to forgive in the South African context (Kaminer et al. 2001). It may be that sex differences in forgiveness pertain to forgiveness of different kinds of violations; indeed, in the setting of the TRC women were more likely than men to report violations to family members (Allan et al. 2006). Speculatively, in the South African setting, forgiving may have higher benefits for males, who are perhaps more involved in commercial interactions with one another than are females. Similarly, in nonhuman primates, it may be hypothesized that reconciliation after contests is particularly adaptive for males who subsequently need to cooperate in key ways (de Waal 2000).

A third point we want to make is that even within groups of individuals with apparently similar interests (from an adaptive perspective), there is considerable variation (consider, for example, Nelson Mandela and Desmond Tutu's emphasis on reconciliation at a time when many of their comrades and colleagues were less concerned with this issue). Indeed, a comprehensive account of the full range of revenge and forgiveness in settings such as the new South Africa would seem to require a model that goes beyond a purely computational approach that weights costs and benefits, to also include an account of how values are embodied in cognitive-affective systems. Thus, for example, a detailed understanding of the moral metaphors that individuals live by, and of underlying motivational structures, appears relevant to a full understanding of the complex range of revenge and forgiveness behaviours seen in real-life settings (Boulding 1969; Johnson 1993).

Along these lines, we would emphasize that our observations and data from South Africa are tangential to the question of whether revenge and forgiveness behaviour is mediated by specially evolved brain "modules." Although tackling the question of whether any specially evolved modules exist is not the main focus of our commentary, we wish to note that, in our view, the brain-mind has evolved in a considerably "messier" way than might be suggested by some views of neat modules proposed by evolutionary psychology, that evolved cognitive-affective systems instead have components with indistinct boundaries and distributed functions, and that behaviour is best understood to emerge from an interaction between relatively few ancient special-purpose circuits and more recent general purpose mechanisms (Nesse & Stein 2012; Panksepp & Panksepp 2000).

In summary, then, based on observations from one country that arguably conducted a nation-wide social experiment on revenge and forgiveness, we would conclude that these are partially consistent with the hypotheses generated by McCullough and

colleagues. At the same time, we note that data were not specifically collected with this thesis in mind, and therefore do not provide a very direct test of the hypotheses. The South African findings on gender are of particular interest insofar as they apparently conflict with other published findings, and yet might be explicable on the basis of particular circumstances in South Africa, and with the underlying explanation that revenge and forgiveness involve adaptive mechanisms, and therefore will be triggered in different ways in different environments. Finally, we suggest that the rich range of revenge and forgiveness phenomena in real-life settings is likely to require explanatory concepts other than specialized "modules" and their computational outputs.

The logic of moral outrage

doi:10.1017/S0140525X12000465

Eric Luis Uhlmann

Management and Human Resources Department, HEC Paris – School of Management, 78351 Jouy-en-Josas, France.

eric.luis.uhlmann@gmail.com

www.socialjudgments.com

Abstract: McCullough et al.'s functionalist model of revenge is highly compatible with the person-centered approach to moral judgment, which emphasizes the adaptive manner in which social perceivers derive character information from moral acts. Evidence includes *act–person dissociations* in which an act is seen as less immoral than a comparison act, yet as a clearer indicator of poor moral character.

In the target article, McCullough et al. propose a functionalist model of revenge in which retaliatory aggression is neither irrational nor a sign of pathological dysfunction. Rather, such behavior is driven by the implicit calculations of a revenge system that seeks to deter harmful future acts against oneself, kin, and allies. This revenge system operates in tandem with a forgiveness system designed to reassess whether a person is worthy of inclusion in one's ingroups.

The authors' analysis of revenge is in harmony with the person-centered approach to moral judgment (Pizarro & Tannenbaum 2011; Tannenbaum et al. 2011; Uhlmann & Zhu, under review; Uhlmann et al., under review a; under review b), which emphasizes the adaptive manner in which social perceivers derive character information from moral acts. In the same spirit as McCullough et al., we argue that what at first appear to be irrational decision-making biases often "make sense" when one considers the adaptive goals individuals must meet as they navigate their social environments. Unlike McCullough et al., we emphasize that people often wish to avenge moral transgressions that not only do not harm them or their kin, but in some cases cause no material harm at all.

In addition to assessing the permissibility of acts, people use behaviors to draw inferences about the moral character of the agents who carry them out. Whereas moral judgments of acts are comparatively more likely to center on the tangible harm caused, judgments of persons focus on whether the behavior signals the presence or absence of positive moral traits (Tannenbaum et al. 2011; Uhlmann et al., under review a; under review b). Relatively harmless acts can therefore provoke outrage when they suggest severe deficits in moral character.

Some behaviors are more informative than others regarding an agent's personal character (Nelson 2005; Nelson et al. 2010; Reeder & Brewer 1979). Drawing inferences about character based on such signals is critical to resolving the collective action problems central to McCullough et al.'s analysis of revenge. Because moral traits predict whether a person will cooperate with us or betray us, character-relevant information becomes extremely valuable when navigating social environments. Even acts that are not especially harmful can speak strongly to personal character and therefore prove useful for anticipating more consequential future acts. For the same reason, relatively harmless acts can elicit a desire to castigate the transgressor and exclude him or her from social ingroups.

Evidence for the person-centered approach to moral judgment is provided by *act–person dissociations* in which an act is seen as less immoral than a comparison act, yet as a clearer indicator of negative moral traits. For example, although beating one's girlfriend is viewed as more morally blameworthy than beating her cat, the latter act is seen as indicating a more coldhearted and sadistic person (Tannenbaum et al. 2011). This might seem like a bias in moral judgment until one recalls that animal cruelty predicts antisocial behaviors and an erosion of normal empathic responses (Becker et al. 2004).

Additional studies demonstrate act-person dissociations in the context of truly harmless acts. For instance, although the use of a racial slur (in private, and with no one overhearing) was seen as a less blameworthy act than physical assault, use of a slur was perceived as providing more negative information about the person's character (Uhlmann et al., under review a). Consistent with the idea that person-centered judgments serve the function of determining who to include in one's social ingroups, participants were more willing to be friends with the target who had been physically aggressive than with the bigot.

Negative gut reactions to harmless-but-disgusting transgressions are frequently cited as a case of moral bias. Indeed, participants can find themselves dumbfounded when asked to justify why they feel eating a dead dog and having sex with a chicken carcass are morally wrong (Haidt 2001; Haidt et al. 1993). Part of the reason for strong intuitive responses to such transgressions is that they provide more diagnostic information about the personal character of the agent than do most harm violations. Participants rated eating a dog to be less immoral than stealing a steak, yet more informative of poor moral character (Uhlmann & Zhu, under review).

People are left at a loss to justify their intuitions regarding harmless-but-disgusting acts because they cannot be defended using rational criteria such as the degree of harm caused. In contrast, because of their high informational value regarding underlying traits, there is a clear rational basis for drawing strong character inferences from such behaviors. Although participants were morally dumbfounded when asked whether sex with a chicken was an immoral *act*, they were not at all dumbfounded when asked whether a *person* who engaged in sex with a chicken had negative moral traits (Uhlmann & Zhu, under review). This lack of dumbfounding regarding person judgments was driven by the behavior's high informational value.

As further evidence that person-centered judgments are not subjectively irrational (Pizarro & Uhlmann 2005), act–person dissociations are observed under conditions of both joint and separate evaluation (Tannenbaum et al. 2011; Uhlmann & Zhu, under review). Joint evaluation, in which social targets are evaluated side-by side, promotes logical comparisons and attenuates many decision-making biases (Gaertner & Dovidio 1986; Hsee et al. 1999). This suggests that participants do not view their tendency to judge actions and persons differently as irrational. If they did, they would correct their judgments under conditions of joint evaluation so as not to show any "bias."

In closing, I agree with McCullough et al. that vengeance against wrongdoers – like moral outrage more generally – has a logic to it. Indeed, their functional analysis of revenge converges with our findings in highlighting the adaptive, reputation-based, and person-centered nature of moral cognition. The desire to exact revenge – and willingness to ultimately forgive – may often be less about the transgression itself than what it says about the agent's moral worth as a person.

A systems view on revenge and forgiveness systems

doi:10.1017/S0140525X12000611

Tyler J. Wereha and Timothy P. Racine

Department of Psychology, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.

tjw4@sfu.ca

tracine@sfu.ca

Abstract: Applying a non-developmental evolutionary metatheory to understanding the evolution of psychological capacities leads to the creation of models that mischaracterize developmental processes, misattribute genes as the source of developmental information, and ignore the myriad developmental and contextual factors involved in human decision-making. Using an evolutionary systems perspective, we argue that revenge and forgiveness cannot be understood apart from the development of foundational human psychological capacities and the contexts under which they develop.

The mechanisms underlying the capacities for revenge and forgiveness have evaded evolutionary analysis despite the direct and indirect fitness consequences these forms of social interaction have. McCullough et al. provide an adaptational account for the evolution of complementary cognitive systems for revenge (deterring future harm) and forgiveness (mitigating possible relationship damage and avoiding the costs associated with such a loss). We argue, however, that the conceptualization of such systems is problematic, and the social exchange research reviewed sheds very little light on the evolution of these purported cognitive systems. The authors base their model on a fundamentally non-developmental evolutionary paradigm that misconstrues the nature of development, its role in evolution, and mischaracterizes the psychological capacities they wish to explain. In the following, we outline how an evolutionary systems approach problematizes the plausibility of such systems and their putative selection and specification. An evolutionary systems perspective requires examining developmental processes, and in this case would require reviewing the social cognitive development research that demonstrates the complex and variable ontogenetic processes that give rise to the highly developed perspectival understanding required in the social exchange scenarios they describe.

Despite giving "equal footing" to proximate factors in their adaptationist account (e.g., appealing to Tinbergen's four questions; target article, sect. 2.1) developmental processes are *a priori* ignored or mischaracterized by the non-developmental evolutionary metatheory involved. McCullough et al. claim that a cost is only revenge if it is "*caused by a mechanism designed to deter cost-impositions or benefit-withholdings in the future*" (sect. 2.2.4, their emphasis). From this evolutionary psychological (EP) perspective, decisions to act are "computations." They are evolutionary-derived algorithms (instantiated as cognitive mechanisms) that were selected because they conferred fitness advantages upon ancestral humans. These algorithms do some heavy lifting in this model, in that they are able to compute courses of action for seemingly countless situations and factors involved.

The question is: To what extent is it useful to consider the decisions humans make in a social exchange game as the product of evolved cognitive mechanisms? Aside from the platitude that we are products of natural selection, can we gain any insight into the specifications of such a system through reverse-engineering our decisions to seek revenge or forgive our transgressors? From the non-developmental population genetic perspective of the authors, this is possible. That is, genes that give rise to particular, adaptive phenotypes were selected for in the past. Because the genotype is taken to be the source of developmental information, it follows that genes represent the entirety of the phenotypic outcomes that were once adaptive and selected to solve recurrent social problems. Environmental factors, at most, become a trigger for the initiation of a developmental program, and are thus deemed "equally important."

The neo-Darwinian paradigm has been increasingly criticized for its neglect and mischaracterization of developmental processes (see e.g., Pigliucci & Müller 2010). Failing to use an evolutionary systems perspective that fully incorporates developmental processes into evolutionary theory leads McCullough et al. to mischaracterize the nature and source of developmental information. For example, McCullough et al. do not seriously consider that the evolutionarily relevant problems they outline not only occurred in our evolutionary past, but they are still occurring. The implication is that these problems are re-occurring *developmental* problems, that are, in part, created anew each generation, and with which each generation must contend. A specific genotype does not equate to a specific phenotype (Charney 2012). Simply stated, genes do not contain the developmental information in the way needed to make McCullough et al.'s model tenable.

From an evolutionary systems perspective, examining a "revenge system" would include tracing the development of the capacities underlying such a system. The complexity of the "system" would constantly change (i.e., develop) over time, with various factors changing it in sometimes non-obvious ways. In the particulate, additive, non-developmental stance popular in EP, the "system" with all its potentialities are developmentally predetermined, a perspective that is contrary to the probabilistic nature of development (Gottlieb 2007). Further, selection only selects products of developmental processes. A selectionist perspective can at best only describe why some abilities persist, not how they appeared, which has little relevance for understanding the choices people make in a social exchange situation.

For McCullough et al., "evolved cognitive mechanisms" underlie decision-making processes. Decisions, however, are made by people with life histories. Decisions are influenced by knowledge and emotional and motivational states; they are a function, in part, of one's experience interacting with others within particular social and cultural norms of conduct based upon forms of interaction characteristic of biological humans. A developmental perspective is required to examine the many foundational psychological capacities that lead to the high level of social understanding of adult humans. This understanding has a long and complex ontogeny beginning in the first year of life with the capacity for joint attention, the ability to coordinate attention and activity with oneself and another. Young children develop a "theory of mind" around age 4 when they are able to understand false belief (Carpendale & Lewis 2006). Although the consistent age of false belief understanding has led many to assume that it is specified by genetic or biological factors (i.e., Baron-Cohen 1995; Gopnik 1996; Leslie 1987), Boesch (2007) demonstrates the extreme variability in false belief understanding across cultures, including those that do not pass a standard false belief test at age 14. Such findings reiterate the fact that consistent development takes place in the context of consistently recurring developmental factors. Little can be said about "evolutionary mechanisms" when the development of these capacities is not examined.

A fundamentally non-developmental evolutionary metatheory is not conducive to understanding the evolution of psychological capacities. From an evolutionary systems perspective, understanding developmental changes over generations is to understand the evolution of a trait or a cognitive system. A selectionist perspective and its inherently developmentally uncontexualized mechanisms does little to provide this understanding. The existence of revenge across cultures does not mean it is universal in the sense that it is based on genetic selection, but results from the common developmental conditions of our species.

Revenge, even though it is not your fault

doi:10.1017/S0140525X12000623

Rongjun Yu

Department of Psychology and Center for Studies of Psychological Application, South China Normal University, Guangzhou, 510631, China.
rongjun.yu@gmail.com **www.rongjunyu.com**

Abstract: McCullough et al. argue that revenge has a future-oriented function, that is, to deter future harms by changing other individuals' incentives toward the self. Recent research has shown that people seek revenge even when harms are unintentional. This commentary reports these results and proposes that revenge may also serve to reduce the immediate psychological pain resulting from unfair treatment.

Two factors play a key role in criminal conviction in the common law tradition: a harmful consequence (*actus reus*) and the intent to harm (*mens rea*). Intentions at the time of action influence moral judgment and the subsequent punishment (Young et al. 2010). Individuals who harm others accidentally and unknowingly receive less punishment than those who harm others intentionally. McCullough et al. propose that revenge functions to deter future harms by increasing others' "welfare tradeoff ratios" (WTRs) toward victims. This theory implies that revenge should exist only when harms are intentional because only intentional harms can reveal others' WTRs. On the other hand, unintentional harms are not informative about the harm doer's true WTR and thus should not invite revenge. However, punishment of innocent people is not uncommon in real life situations. Take envy, for example; disliking others' wealth leads people to pay to destroy the envied person's money, even though the envied person is not responsible for the inequality. In laboratory studies, evidence suggests that revenge exists even when harms are unintentional.

Our recent functional magnetic resonance imaging (fMRI) study investigated the behavioral and neural responses to different types of fairness (Yu et al., submitted). In our experiment (see Fig. 1), two participants (strangers) jointly completed a matching task and then they received the outcomes. If their choices were matched, they both received some monetary rewards. Otherwise, they both lost money. However, the exact amounts of money each player could win or lose in each trial were determined by a computer program. Participants received advantageous (more than their partner), disadvantageous (less than their partner), or equal payoffs. Then, they were given the opportunity to alter their partner's payoff at their personal costs. It is a one way punishment, that is, their partners did not have such opportunity to punish. Every fifty pence increase or decrease in the partner's payoff cost participants ten pence.

We found that individuals have strong preferences for fairness in both disadvantageous and advantageous inequality conditions, such that they alter others' payoff toward an equal distribution at a personal financial cost. At the neural level, individuals who spent more money to increase others' payoff had stronger activity in the putamen (the reward region) when they encountered advantageous inequality (Mobbs et al. 2009). Conversely, those who spent more money to reduce others' payoff had stronger activity in the amygdala (the anger region) in response to disadvantageous inequality, suggesting that negative emotions evoke revenge (Scott et al. 1997). Revenge may reduce the immediate psychological harms (e.g., envy and anger) by bringing others down. Our study suggests that accidental harms are enough to elicit immediate negative emotions which may evoke the desire for revenge.

Why do people punish those who are not responsible for the inequality? One possibility is that when an individual is treated unfairly, the induced negative emotion is quite intense. Like physical pain, which makes people punch objects, and frustration,

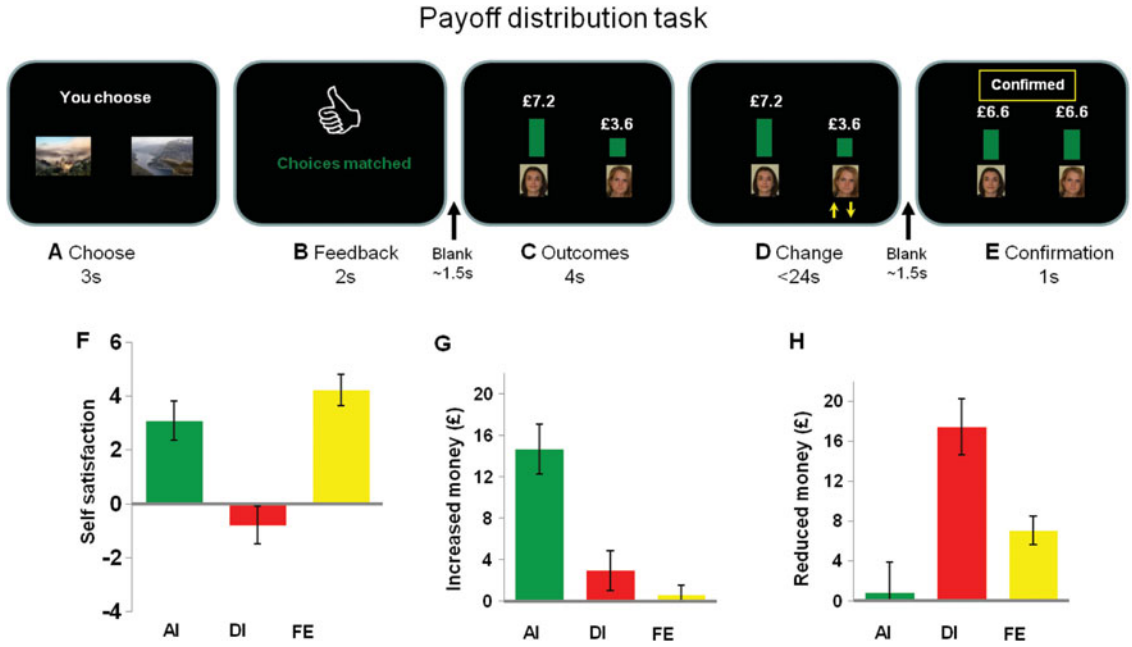

Figure 1 (Yu). **Experimental task design and behavioral results.** (A) In the payoffs distribution task, participants were required to choose one image. (B) After the Choose stage, participants were informed whether their choices were matched or not, and hence, both win or both lose. (C) The outcome for the participants and the outcome for their partners were presented. (D) After the Outcome stage, participants could alter the partner's payoff at their personal costs. (E) Participants pressed a third key when they finished changing. The final payoffs for both players were depicted. (F) The self-reported satisfaction toward outcomes across win and loss trails in advantageous inequality condition (AI), disadvantageous inequality condition (DI), and fair equal condition (FE). (G) The increased money (total money spent to increase other's payoffs) in each condition. (H) Reduced money (total money spent to decrease other's payoffs) in each condition.

which provokes displaced aggression, the social pain resulting from inequality drives people to revenge. The psychological urge to reduce immediate pain is ignored in the target article. From an evolutionary perspective, the sense of unfairness is vital for an individual's survival in social situations and thus revenge may have evolved as an instinctive reaction to unfair treatment. Furthermore, outcomes are easy to evaluate but intentions are difficult to know. Negative outcomes may be enough to elicit revenge motives in the initial stage. Whether to take revenge or to forgive is modulated by attribution of intentions in the latter stage. Previous research shows that forgiveness requires the effort to restrain vengeful impulses (DeWall et al. 2007; 2010), suggesting that revenge is an emotional "hot" system and forgiveness is a rational "cool" process.

Revenge may ultimately hurt the seeker as much as the victim (Dreber et al. 2008). Like a proverb states, "Before you embark on a journey of revenge, dig two graves." In our study, punishment reduces both players' payoffs and participants know that. Even when revenge seekers know such consequences, for example, for individuals who commit crime of passion, they still choose to do so. I argue that revenge is not always future-oriented and may have evolved for other reasons, such as the fairness instinct. It is possible that revenge serves to restore the fairness social norm when individual self-interest has been violated by others. It functions mainly to reduce current emotional harms rather than to deter future harms. This explains why in many situations people seek revenge even when it escalates conflicts rather than moderates them, leading to destructive outcomes for everybody involved.

Authors' Response

doi:10.1017/S0140525X12001513

Putting revenge and forgiveness in an evolutionary context

Michael E. McCullough,[a] Robert Kurzban,[b,c] and Benjamin A. Tabak[d]

[a]*Department of Psychology, University of Miami, Coral Gables, FL 33124-0751;* [b]*Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104;* [c]*Economic Science Institute, Chapman University, Orange, CA 92866;* [d]*Department of Psychology, University of California–Los Angeles, Los Angeles, CA 90095-1563.*

mikem@miami.edu
http://www.psy.miami.edu/faculty/mmccullough
kurzban@psych.upenn.edu
http://www.psych.upenn.edu/~kurzban/
btabak@psych.ucla.edu

Abstract: In this response, we address eight issues concerning our proposal that human minds contain adaptations for revenge and forgiveness. Specifically, we discuss (a) the inferences that are and are not licensed by patterns of contemporary behavioral data in the context of the adaptationist approach; (b) the theoretical pitfalls of conflating proximate and ultimate causation; (c) the role of development in the production of adaptations; (d) the implications of proposing that the brain's cognitive systems are fundamentally computational in nature; (e) our preferred method for considering the role of individual differences in computational systems; (f) applications of our proposal to understanding conflicts between groups; (g) the possible implications of our views for understanding the operation of contemporary criminal justice systems; and (h) the question of whether people ever "genuinely" forgive.

R1. Introduction

We are grateful to the many scholars who took the time to read and consider our target article. Despite their potential importance to social life, revenge and forgiveness have been, we think, undertheorized (McCullough 2008), and it was our hope that through an adaptationist analysis of behavior and a computational understanding of cognition we might help to stimulate the sorts of research projects in the future that would contribute to a fruitful consilience of the social, behavioral, and life sciences (E. O. Wilson 1998). As **Konečni** points out, the scientific record is full of important empirical results that are relevant to our claims, although inevitably we failed to find all of them. We are thankful for those that commentators such as Konečni have brought to our attention.

The commentators have raised issues of two broad types: first, those that concern the specific claims about revenge and forgiveness that emerged from our approach, and, second, those that concern the meta-theoretical apparatus we put to work in our analysis. We broke responses down further into eight substantive themes. In this response we take the eight themes in turn. In Section R2 we discuss the inferences that are and are not licensed by patterns of contemporary behavioral data in the context of the adaptationist approach. In Section R3 we describe the theoretical pitfalls of conflating proximate and ultimate causation. In Section R4 we clarify our stance on the role of development in the assembly of adaptations. In Section R5 we lay out the implications of proposing that the brain's cognitive systems are fundamentally computational in nature. In Section R6 we describe our approach to considering the role of individual differences in computational systems. In Section R7 we comment on the possible applications of our theorizing to conflicts between groups. In Section R8, we explore the possible implications of our views for understanding the operation of contemporary criminal justice systems. Finally, in Section R9 we consider the question of whether people ever genuinely forgive.

R2. What are the entailments of claiming that cognitive mechanisms for revenge and forgiveness are adaptations with identifiable functions?

Some commentators believe our analysis of revenge and forgiveness leads to implausible hypotheses about the widespread occurrence of revenge in human societies (**Gintis**), that the analysis "sheds very little light on the evolution of these purported cognitive systems" (**Wereha & Racine**), and that adaptations for revenge and forgiveness are unlikely to exist at all (**Holbrook, Fessler, & Gervais** [**Holbrook et al.**]). **Barclay** is right in pointing out that one major risk of adaptationist analysis is that readers might misperceive functional claims as universal claims. The claim that *the revenge system is an adaptation* emphatically does not entail that *all instances of revenge (or forgiveness) will be adaptive* (Andrews et al. 2002; West-Eberhard 1992; Williams 1966).

Gintis bases his skepticism of our central claims, which he calls "implausible," on (a) evidence from economics experiments indicating that third parties will, under some laboratory conditions, pay costs to punish a stranger who has been stingy or greedy with regard to another stranger;

(b) educated guesses about ancestral population structure; and (c) the results of his and his colleagues' evolutionary simulations. Gintis's claims, and the evidence he adduces in support of them, have recently been addressed extensively elsewhere (e.g., Guala 2012; West et al. 2011), so we restrict our comments to Gintis's first point, which bears most strongly on our central claims.

Although the data from economics experiments used to support the concept of strong reciprocity are interesting and important (but see sect. 3.1.3 in the target article for difficulties surrounding interpretation), the finding that people will sometimes pay costs to punish harmdoers even when (by experimental constraint) the punishers cannot benefit economically or reputationally from doing so does not damage our claim that humans have adaptations for punishment that were designed by natural selection because of their deterrent effects. Zooming out to a broad conceptual level, adaptationists since Williams (1966) have hewed to a definition of an adaptation as "a characteristic of an organism whose form is the result of selection in a particular functional context" (West-Eberhard 1992, p. 12). A critical entailment of this definition is that a trait's status as an adaptation must be evaluated from the perspective of the *historical selection pressures that gave rise to the trait's gene-propagating effect*. The operation of the adaptation within the organism's contemporary ecology (or within a laboratory experiment) might accurately reflect the function that natural selection designed the adaptation to perform, but to the extent that the organism's contemporary ecology (or the laboratory experiment) fails to capture key elements of the ecological backdrop against which natural selection gave rise to the adaptation in question, contemporary results from both the field and from the laboratory can be deeply misleading (Burnham 2003; Hagen & Hammerstein 2006). Indeed, an adaptation can appear to be "misfiring," even though it is merely executing its proper function in response to environmental stimuli whose sensory properties are close enough to approximate the ancestral conditions under which the adaptation was naturally selected to operate.

Biologists frequently encounter initially puzzling costly contemporary behaviors. In Colorado, for example, yellow-bellied marmots have not encountered wolves since the 1930s, when farmers and ranchers eradicated them. Nevertheless, when exposed to life-sized two-dimensional images of wolves during field experiments, these marmots immediately suspend their foraging activity to run and hide – a costly pattern of behavior that is unlike their responses to equivalent images of extant predators or control animals (Blumstein et al. 2009). Costly fleeing in marmots, therefore, can be explained with reference to benefits that existed in the past but no longer do. The fact that marmots receive no benefits in the modern ecology from fleeing from images of wolves does not make implausible, as **Gintis**'s argument would have it, that the mechanisms for fleeing were selected for by virtue of the fitness benefits those behaviors used to provide (viz., avoiding predation by wolves) under ancestral circumstances.

Therefore, for some stimulus-response relationships, it is *more* plausible that the adaptation that causes a given behavior is firing in response to a stimulus that is outside of the adaptation's proper domain (i.e., the range of stimuli whose biological function it is to process; see Sperber 1994), than it is that the adaptation's proper domain is broader than scientists had previously apprehended. To the extent that in ancestral human environments repeat interactions were common (Hagen & Hammerstein 2006), cooperation-regulation mechanisms might embody the *ex ante* assumption that interactions are likely to be repeated (Delton et al. 2011) even though some social interactions might have turned out (*ex post*) to be one-shot. And if this were the case, people in modern (e.g., laboratory) environments should be expected to execute behaviors that promote cooperation or deter exploitation (including punishment) even when they are aware (i.e., have an explicit representation) that the interaction is likely to be one-shot. Thus, for some of the same reasons why misfiring arguments are better suited to explaining some aspects of contemporary marmot behavior, we also think that misfiring interpretations are better suited than is **Gintis**'s account for explaining the existing experimental results about "strong reciprocity."

R3. What are the advantages of carefully distinguishing between ultimate and proximate levels of causation?

Several commentators seem unconvinced that our model presented an exhaustive account of the causal forces that are operative within revenge and forgiveness systems. **Potegal**, for example, claims that we fail to give adequate weight to the "reinforcing value of aggression" as part of the causal apparatus that makes revenge happen. Other commentators (e.g., **Fatfouta, Jacobs, & Merkl** [**Fatfouta et al.**]; **Dellis & Spurrett**; **Yu**; **Ross**) have suggested in one way or another that our analysis gave insufficient attention to neurological evidence (or the lack of neurological evidence) about the systems that might be involved in the production of revenge and forgiveness.

Some of these misgivings, we think, are traceable to confusion about the differences between ultimate and proximate causation (Scott-Philips 2008). In the target article, we took pains to point out in sections 2.1 and 2.2 that explanations for revenge that are based on statements about proximate causation (e.g., that people enact revenge because it feels "satisfying") are inadequate for explaining the evolution of such mechanisms in the first place. Ultimate causal explanations must always be with respect to the fitness-enhancing (i.e., gene-propagating) effects of rival designs (Scott-Philips 2008). So, when **Ross** suggests that "a key condition for the existence of equilibrium strategy vectors that include revenge and forgiveness" (by which we take him to mean that the conditions under which revenge and forgiveness can become evolutionarily stable strategies; Smith 1982) is the human susceptibility to shame, we believe that more careful attention should be paid to the proximate/ultimate distinction: Shame, which is an emotion, is a proximate causal force (which, one is free to argue, is put to use by cognitive systems for revenge and forgiveness) rather than a statement about the effects of rival designs on gene frequencies (Scott-Philips 2008). Consequently, shame cannot be invoked to describe the ultimate casual forces that lead to the assembly of mechanisms for revenge and forgiveness. A similar conflation of ultimate and proximate levels of causation occurs

when **Yu** writes that "Revenge is not always future-oriented and may have evolved for other reasons, such as the fairness instinct." **Sell**'s commentary is an elegant example of how considering ancestral selection pressures (viz., for up-regulating a harmdoer's WTR for oneself versus reducing someone's capacity for harming oneself) can yield fine-grained predictions about the distinct proximate characteristics of the psychological systems that generate anger versus hatred.

R4. Does our theorizing ignore the role of development?

Wereha & Racine take us to task for using a "fundamentally non-developmental evolutionary paradigm," and commend as an alternative their own "evolutionary systems perspective." Their critique gets its traction, however, by assigning views to us that we do not endorse, and by substituting a set of bold but largely empty propositions about understanding the development of the adult revenge system. Most flagrantly, they claim that, "In the particulate, additive, non-developmental stance popular in EP [evolutionary psychology], the 'system' with all its potentialities are developmentally predetermined, a perspective that is contrary to the probabilistic nature of development." Many researchers in evolutionary psychology, however – including those whose views we would associate with our own – take development very seriously indeed (Belsky et al. 1991; Ellis 2004; Ellis & Bjorklund 2005; Frankenhuis & Del Giudice 2012; Geary & Bjorklund 2000), and have taken pains to distance themselves from the particulate, additive, adevelopmental caricature that Wereha & Racine set up as representative of evolutionary psychology's stance on development (Tooby et al. 2003). Tooby et al. (2003) pointed out that the idea that development is a complex interaction between genes and environment is the *starting point* for evolutionary psychologists.

Wereha & Racine's critique is further undermined by a category error. Seeking to contrast our view that "'evolved cognitive mechanisms underlie decision making processes" (their phrasing), they write that "Decisions, however, are made by people with life histories." Cognitive mechanisms at any given moment both *cause decisions* and *have developed over time*, so setting the two in opposition to each other is illogical. Further, our colleagues in evolutionary psychology who study development draw heavily from life history theory (Ellis 2004), a feature that distinguishes their approach from other developmental approaches, thereby hollowing out this critique. Although it is true we did not focus on development in the target article, we do not think that the revenge system magically appears in adult form; indeed, in section 3.3.2, we discussed the role that people's life histories can play in altering the operation of mechanisms for revenge (see also **Barclay)**. We take for granted that the unique elements of people's life histories likewise influence the operation of mechanisms for forgiveness. Similarly, the idea that we "do not seriously consider" that people still face the problem of deterrence is a striking misrepresentation of our work. We discuss examples from the lab and the field in which people are faced with the category of problem that we believe selected the behavior, which is that aggression now predicts aggression later.

Finally, we note that "systems" theories have been criticized for yielding only predictions that are vague at best. Indeed, Tooby et al. (2003) suggested that "developmental systems theory makes no predictions." Vindicating this bold assertion, the closest **Wereha & Racine** come to a positive statement about what their view predicts is the claim that the "complexity of the 'system' would constantly change (i.e. develop) over time, with various factors changing it in sometimes non-obvious ways." It is difficult to imagine what pattern of empirical data might put such a claim in jeopardy.

Generally, we emphatically agree that explaining and understanding the development of revenge systems is an important priority. Indeed, work such as Sell et al.'s (2009) research on anger points to a potentially profitable direction: identifying factors – in Sell et al.'s case, variation in size and attractiveness – that might be expected to systematically influence developmental outcomes. Hypotheses about the privileged roles of ecological factors such as local life expectancies, frequencies of within-group interpersonal violence and intergroup warfare, the strength of fraternal interest groups, and the harshness of one's family environment during early childhood likewise merit exploration in future developmental work on revenge and forgiveness.

R5. What does it mean to refer to systems for revenge and forgiveness as computational? What does it mean to refer to them as *systems*?

Several commentators express reservations about our claim that the mechanisms underlying revenge and forgiveness are computational systems. More specifically, some argue that it is important to consider that these systems might not be "rational" (**O'Connor & Adams**) as opposed to emotional (**Aureli & Schaffner**; **Leiser & Joskowicz-Jabloner**), and that our explanation was either, on the one hand, unnecessarily complex (Aureli & Schaffner) or, on the other, insufficiently so (**Stein, van Honk, & Ellis** [**Stein et al.**]).

First, we wish to clarify that we were not trying to be tendentious in making a computational claim. Following convention in the cognitive sciences (e.g., Carruthers 2006; Pinker 1999), we take computation to be the information-processing description of what the brain's functions entail. Computational mechanisms take as inputs select types of information (including possibly information from other computational systems), represent that information in some sort of physical format, perform operations on those representations, and pass the outputs of those operations to other neural or somatic systems for further processing or action production. So, in our view, systems for revenge and forgiveness are computational because their function is to represent and process particular types of information – specifically, information that would have led them to cause good (i.e., fitness-raising) decisions in the domains of the adaptive problems for which they were naturally selected (i.e., deterring future harms and updating aggressors's intrinsic WTRs for the self peaceably).

We hope it goes without saying that we take for granted that natural selection is the only cause of complex functional design in biology – computational design included. Neural systems can only be called computational to the

extent that they physically represent states of the world with a non-zero degree of fidelity; thus, to the extent that computational systems are beset by mutations or otherwise influenced by other non-adaptive causal process, the effect of these non-adaptive causal processes will generally be to reduce the fidelity with which the systems can represent true states of affairs. Consequently, their computational powers will be reduced. For this reason, the concepts of adaptation and computation tend to go hand in hand. To the extent revenge and forgiveness systems exist, we assume that it is fitting to conceptualize them as computational systems, and that "good computation" within their respective domains is shorthand for *computations that ancestrally would have provided reasonable tradeoffs between the benefits of deterrence to be gained by imposing a retaliatory harm on the harmdoer and the relationship-mediated benefits to be gained by signaling one's willingness to withhold revenge and return to mutually beneficial relating, conditional on better treatment from the harmdoer in the future* (Burnette et al. 2012).

Referring to revenge and forgiveness systems as computational systems explicitly is useful, we think, because it keeps one mindful of the need for clear information-processing specifications when investigating how these systems might perform their tasks. On the basis of computational reasoning, for example, Burnette et al. (2012) made five novel predictions about the cognition of individuals who are actively making decisions about whether to forgive or avenge a recent harm:

> (a) those individuals should be willing to pay a relatively large cost to obtain information that is relevant to assessing relationship value and exploitation risk (in comparison to the prices they would pay for other types of social information about the harmdoer), (b) such information should gain privileged access to attention and working memory and should be relatively resistant to interference from competing information, (c) such information should be automatically scanned to determine whether it is the result of deception on the part of the exploitive individual, (d) memories about the exploitive individual that are retrieved from episodic memory should tend to be (on average) valid for evaluating those individuals' relationship value and exploitation risk, and (e) memories about exploitive individuals' past behavior toward the self should be given more weight in decision making than will cues about their behavior toward other individuals. (pp. 353–54).

These predictions resulted from applying computational thinking to how a well-designed forgiveness system might operate, as we tried to do in the target article.

The emotions such as those alluded to by the commentators as potential *alternatives* to the computational steps that might be involved in motivating revenge and forgiveness (e.g., anger), by our reckoning, *are also* computational in nature (see, e.g., Tooby & Cosmides 2008). Indeed, we are unsure what else the systems that produce emotions might be, though admittedly emotions seem special because they are associated with complex conscious experiences in a way that non-emotional cognitive processes (e.g., vision) are not. The question is not whether anger plays a role in the production of revenge (it certainly seems like it does – perhaps along with hatred, as **Sell** perceptively proposes), but, rather, what the computational processes are that make anger happen, and how those subroutines lead to outputs that can then be put to use in the production of revenge. In the target article, we intentionally refrained from implicating anger and similar emotions as causal elements in the information processing stream that leads motivation for revenge and forgiveness because our goal was to articulate the computations these systems must perform, but this was not a move on our part that was designed to exclude emotions in any sense. So, we find little reason to quarrel with **Aureli & Schaffner**'s suggestion that "emotional mediation" is important for revenge and forgiveness: The computations involved in the production of revenge and forgiveness no doubt involve emotions; emotions are, from our point of view, fundamentally computational in nature, too.

Further, our claim was not that the computational systems we propose are "rational" as economists use the term. "Anger" might be well designed to deter, as we have proposed, yet give rise to "irrational" behavior, such as (vengeful) rejections of low offers in the Ultimatum Game. According to this view, emotional systems execute their evolved function in a way that respects ancestral computations of costs and benefits, but far from the way envisioned by standard bloodless economic analysis (Frank 1988). So, (emotional) deterrence systems need not be "deliberative," and we reject as ill-formed the persistent "tension between affect and deliberation" to which **O'Connor & Adams** allude (Tooby & Cosmides 2008).

As for the complexity required for such systems, we did not intend to take a strong stand on this, though would resist the flavor of **Aureli & Schaffner**'s remarks that emotional systems are necessarily simpler than non-emotional systems. In our view, emotional systems embody potentially intricate complex computations (Tooby & Cosmides 2008), and we look forward to work from people from differing perspectives helping to shed light on the intricacies of the involved computations. For example, **Johnson-Freyd & Freyd** point out that one way to engineer "acceptance" is to engineer the systems to ignore – or at least appear to ignore – intentional harms. We agree that ignorance can have strategic advantages (Kurzban 2010b; Schelling 1960), and we are sympathetic to the view that one means of accepting an offense, in terms of the outcome for the relationship, is to ignore or pretend to ignore an offense.

Holbrook et al. are not sanguine about the claim that human brains contain computational systems whose function is to deter future harms. Comparative data illustrate that nonhuman animals from multiple taxa impose retaliatory harms on other organisms that have previously harmed them, and that by doing so, the retaliators deter the recipients of their retaliatory behaviors from harming them again in the future (e.g., Aureli et al. 1992; A. Bshary & Bshary 2010; R. Bshary & Grutter 2005; Hoover & Robinson 2007; Jensen et al. 2007). Such data make the parallel claim for humans plausible.

Instead of revenge mechanisms designed for deterrence, **Holbrook et al.** propose as an alternative hypothesis that the deterrence function we have in mind is more parsimoniously handled by anger along with a variety of domain-general systems (e.g., norm acquisition, future forecasting, and perspective taking) and "systems related to other motivations, such as reputation management." However, we would argue that their proposal that revenge is subserved by "evolved capacities to categorize events, assume

others' perspectives, forecast the future, and weigh costs against benefits" lacks substantial theoretical or predictive force because it constitutes a far too general gloss of a computational system. The capacities that Holbrook et al. catalogue are all, to be sure, important in executing a deterrence function, but the organism that is effective at deterring others from imposing harms upon it in the future must be motivated – by, for instance, the experience of anger – to take appropriate adaptive action out of all the possible actions that one might take. Categorizing events and so on is also insufficient; particular events (intentional harm) need to be met with appropriate behavior (e.g., return harm) to bring about adaptive outcomes. The suggestion that people choose how to react to a situation through categorizing it and forecasting the future allows no predictive mapping between situation and behavior, specifying only the sorts of mechanisms that are recruited as opposed to what strategies those mechanisms ought to implement.

To put it another way, the claim that there is no deterrence system *per se* implies that to the extent that people's propensity to harm in response to harm does deter, this comes about as an incidental side-effect of the action of systems designed for some other (perhaps more general) function. What function might anger have, such that it is not designed to deter but happens, as a lucky side-effect, to deter? The answer to this cannot be "reputation management" without a more explicit and specific account of precisely how one ought to manage one's reputation. Why not cultivate a reputation to be unmoved, or even happy, when one is harmed? There is an arbitrarily large vocabulary of reputations one might cultivate; one cannot cultivate a "good" reputation unless one specifies the problem that having a reputation is supposed to solve. Our posited deterrence function explains why people experience anger rather than joy at being harmed; a "reputation management" function, in itself, does not (see Tooby & Cosmides 1992, pp. 109–13; 2008).

So, even if it is the case that the devices that are wired together to generate revenge within human brains are also used for other functions (as **Holbrook et al.** posit), then an explanation is required for how these devices came to be wired together in just the sort of bricolage, to use their favored term, that causes that bricolage to create retaliatory behavior that returns fitness-enhancing deterrence benefits to its bearer. Neural wiring is expensive and needs an explanation in terms of natural selection every bit as much as do the structures that get wired together (Anderson 2010; Cherniak et al. 2004).

Finally, we respectfully disagree with **Holbrook et al.**'s claim that we have reified a folk category, a claim which seems to rely wholly on their observation that "[t]here are many kinds of deterrence that do not stem from the anger-hatred nexus." In fact, we never claimed that revenge was the only system that deters, and their example of swatting a begging dog to deter future begging shows only that some deterrence is not revenge. The poisons of poison frogs and the thorns of roses are also deterrence mechanisms, for instance. Therefore, we of course agree that there are other systems beyond the ones we posit that can be deployed for deterring predators from attacking or, perhaps, training domesticated animals. More generally, we are comfortable with the notion that the revenge system, as a whole, makes use of subsystems such as the ones that Holbrook et al. identify.

Dellis & Spurrett have a different problem with our specification of the systems involved in revenge and forgiveness: They suggest that there is no reason to postulate two separate systems when one "reciprocity" system will do, suggesting that the key dispositive evidence would be a dissociation due to, for instance, "local brain damage" or "genetic intervention." To clarify our position, we believe that evidence for a putative function can be aided by these empirical patterns, though behavioral evidence of special design is similarly of value (Andrews et al. 2002; West-Eberhard 1992; Williams 1966). We similarly believe mechanisms ought to be (and generally are) individuated by virtue of their function (Barrett & Kurzban 2006). Therefore, we think that it will continue to be useful, in guiding empirical research, to distinguish the function of deterring others from harming oneself, from the function of encouraging others to deliver benefits to oneself.

In short, in broad strokes, we are very pleased to agree with **Aureli & Schaffner**, as well as with **Leiser & Joskowitz-Jabloner**, **O'Connor & Adams**, and **Holbrook et al.** that emotions do important jobs in the production of revenge and forgiveness; but our view is that emotions are computational entities, and if they are to be deeply understood, the function of the computations they execute should be made explicit in order to make good predictions about the nature and details of these computations.

R6. Did we neglect important individual differences?

Several commentators, including **Potegal**, **Yu**, **Fatfouta et al.**, **Balliet & Pronk**, **Karremans & van der Wal**, **Konrath & Cheung**, and **Roberts & Murray**, feel that we gave inadequate attention to important individual differences. Karremans & van der Wal, for instance, call self-control one of the "basic processes that lead to forgiveness," and worry that its neglect might limit our model's ability to explain "how revenge and forgiveness actually occur." Many individual differences are related to variation in people's propensities to forgive and to seek revenge, including empathy and narcissism, as Konrath & Cheung point out, as well as agreeableness (or social concern; see Balliet & Pronk), conscientiousness, self-esteem, and religiosity (Balliet 2010; Fehr et al. 2010; Hoyt et al. 2005; McCullough 2001; McCullough & Hoyt 2002; McCullough & Worthington 1999; McCullough et al. 2001; Riek & Mania 2012). We are not overly concerned about our model's failure to specify all individual differences that are associated with forgiveness and revenge. Nevertheless, we do look forward to more work in this area by people with expertise in individual differences – perhaps explicitly incorporating some of the common theoretical tools that evolutionary biologists use to make sense of individual differences (Buss 2009; Nettle 2006; 2009; Tooby & Cosmides 1990a).

Our interest in writing the target article was to outline the set of computations that function to deter (the proprietary computations of the revenge system) and to select behaviors that inhibit the revenge system and signal a willingness to re-establish positive relations, contingent on improved behavior from the harmdoer (the proprietary computations of the forgiveness system). We grant that the inputs to and outputs from the proposed revenge

system and forgiveness system can be modified by other systems. However, the mechanisms that act on those inputs before they enter the revenge or forgiveness system, and the mechanisms that act on them after they have left the system for further processing, are not necessarily "basic" (though no less interesting as a consequence).

Narcissism, which **Konrath & Cheung** encourage us to consider, for instance, might indeed influence one's proneness to respond to harms with retaliation rather than forgiveness (Riek & Mania 2011), but this does not make narcissism an individual difference that must be incorporated into the computational theory per se. (It also does not make narcissism a "bias": We see little reason to discard the hypothesis that narcissistic people simply believe that they are entitled to better treatment from everybody [i.e., expect others to hold high WTRs for them] because narcissistically entitled people, it appears, really are both physically stronger [Sell et al. 2009] and better looking [Holtzman & Strube 2010; Sell et al. 2009]). As long as narcissism's associations with revenge and/or forgiveness are due to its influence on computational events taking place outside of the revenge or forgiveness systems, it doesn't necessarily require specification within the computational architecture itself. **Barclay** also helpfully describes how individual differences in people's perceptions of others' actions will change the information that computational systems for revenge and forgiveness end up processing, consequently changing their behavioral outputs.

Another individual difference that emerges repeatedly in the commentaries is "self-control" or "executive control" (e.g., **Yu**; **Fatfouta et al.**; **Balliet & Pronk**; and **Karremans & van der Wal**). Impulsivity (which is often pejoratively called "low self-control" or "poor self-control") might be caused in part by mechanisms whose function is to motivate organisms to exploit local opportunities (i.e., to capture resources, or to avert bad outcomes) whose payoffs can be realized on short time horizons or acceptably high probabilities of attainment (Daly & Wilson 2005). In other words, humans might discount the value of rewards by the time required to wait for them or by the certainty with which they will come to fruition when that wait time has ended (Ballard & Knutson 2009). If the deterrence benefits of revenge can be realized sooner, or with greater certainty, than the relationship-restoration benefits of forgiveness, then natural selection might lead to the evolution of a psychology that, ceteris paribus, produces stronger motivation to execute revenge than to execute forgiveness.

In an alternative specification, one might propose that mechanisms for revenge (as is the case with, say, mechanisms that "like" ice cream) themselves have lower discount rates than do mechanisms for forgiveness (as would be the case with mechanisms that "like" to go on diets). After all, there are good reasons to think that a propensity for action now, or for taking a sure bet over a more speculative one, will result, effectively, in the inhibition of mechanisms that are designed to motivate for restoring relationship-bearing benefits: Generally, reaping the fitness advantages of reciprocal cooperation, for example, seems to require that the benefits, which are realized only over a relatively long temporal horizon, not be discounted too steeply (Stevens et al. 2005). Additionally, individual differences in temporal discount rates (i.e., the rates at which people downgrade the subjective value of future rewards as a function of the time until their receipt) are negatively associated with cooperation during the iterated prisoner's dilemma and similar social dilemmas (Curry et al. 2008; Yi et al. 2007).

A plausible hypothesis, then, about the associations of individual differences in "self-control" with revenge and forgiveness, is that dispositionally impulsive people are more vengeful and less forgiving not because, as **Karremans & van der Wal** suggest, their "executive control resources are…depleted" (a view that is, we acknowledge, highly influential in social psychology, but theoretically and perhaps also empirically problematic; Kurzban 2010a; Molden et al., in press; Navon 1984), but, rather because humans have cognitive mechanisms that weight the value of particular courses of action by the probability that the payoffs associated with those courses of action will be realized (see also **Barclay**, who describes how misperceptions of the costs and benefits associated with revenge and forgiveness can lead to individual differences in revenge and forgiveness broadly). When these valuation-assigning mechanisms estimate that one's time horizon is short, they may motivate courses of action (viz., revenge) whose payoffs (viz., direct or indirect deterrence) can be realized over shorter time horizons (McCullough et al. 2012).

Such a hypothesis finds precedent in work suggesting that people from homes in which nurturance, discipline, and parental care were inconsistent, or from neighborhoods in which violence and economic disadvantage were high, engage in more impulsive and risky behavior as young adults (Griskevicius et al. 2011; Hill et al. 2008). By extension, some of the ecological factors that we discussed in section 3.3.2 of the target article could influence people's propensity for revenge because those factors might ancestrally have been correlated with a shorter average lifespan, thereby creating a selection pressure for psychological mechanisms that can use those cues as inputs to regulate impulsive decision-making.

There is an interesting developmental angle here, which might please **Wereha & Racine**: The impulsive behavior associated with adolescence and young adulthood is often viewed pejoratively, as if there were something wrong with the adolescent brain that causes it to make dysfunctionally impulsive choices (Gullo & Dawe 2008; Hill & Chow 2006). However, it is important to bear in mind that every single one of our hominid ancestors successfully negotiated the challenges of adolescence. Thus, it is highly improbable that adolescent impulsivity is caused by something gone wrong in the brains of adolescents, as such frailties would have been subjected to millions of years of negative selection before they could have become part of the species-typical brain design (Daly & Wilson 2005). Accordingly, we assume that youthful impulsivity is caused by mechanisms that "know," thanks to the operation of natural selection, that the only way to improve fitness when one has not yet reproduced is to increase one's reproductive prospects – which, for adolescents, ancestrally came largely from inter-sex competition (i.e., fighting and displays of mate value) and efforts to manipulate opposite-sex choice by demonstrating one's possession of desirable mate characteristics – that is, by showing off (Geary 2010). In this light, perhaps it is no surprise that the modal homicide among non-relatives is perpetrated by a young, unmarried male against another young, unmarried

male during an "escalated showing-off dispute" or as "retaliation for previous verbal or physical abuse" (Wilson & Daly 1985).

R7. Is our theorizing applicable to revenge and forgiveness between groups?

Do the deterrence model of revenge, and the "social-benefit capture" model of forgiveness, apply to groups as well as to individuals? Certainly **Beckerman**'s as well as **Stein et al.**'s remarks along these lines are intriguing, and the presence of revenge in diverse societies during the course of human history and cross-culturally represents a set of phenomena well worth pursuing.

Still, as **Pietraszewski** suggests, the strategic dynamics get considerably more complex when we move from considering only dyadic interactions to multiple agents, with the case of even just three already introducing considerable nuance and texture. It is for this reason that we worry that the leap from individuals to groups, as intimated by **Crisp & Meleady**, might not be completely straightforward.

Specifically, **Crisp & Meleady** contend that "systems for regulating intra- and intergroup behavior should be intimately linked because they involve precisely the same computational requirements." However, we would be cautious about such a claim. In particular, moving from dyads to groups raises several additional issues. First, even if a third party represents himself or herself as belonging to the same "group" as a person who was harmed, the third party still must compute the extent to which it is worth the costs of taking revenge to deter subsequent harms that might later be directed toward the third party as a consequence of the shared group membership with the victim. Just to take one of many possible complications this raises, the third party must infer whether the harm to his or her fellow group member was (a) because of their group membership, in which case revenge might be worth the cost because the third party could be the next victim, or (b) due to some other cause, in which case taking revenge on behalf of one's group member may be less profitable.

This raises the further complication **Crisp & Meleady** point to, the usual issue of free riding. It might be that I indeed would benefit from taking revenge when a member of my group is harmed, but I would be better off still (if no additional benefit from being a punisher can be expected) if someone else from my group bore the costs of revenge rather than bearing them myself. These factors imply that the computations involved in revenge in group contexts might be different from those in the more simply dyadic case. We note, in this regard, that the literature cited by Crisp & Meleady showing punishment in Public Goods games has a somewhat ambiguous interpretation, for the reasons discussed in section 3.1.3.

As **Pietraszewski** points out, computational requirements increase substantially in complexity as we move even the single step from dyads to triads, and, as he indicates, the proximate systems are indeed likely to be "interestingly different." Because humans can, unlike most other species, form relatively long-lasting, non-kin based alliances, coordinating activities in the service of cooperative, potentially antagonistic activities (Harcourt & de Waal 1992; Kurzban et al. 2001; Tiger 1969), we face strategic complications surrounding building and maintaining alliances (DeScioli & Kurzban 2009a; 2009b; 2012), which includes keeping track of where one stands in others' alliance hierarchies (DeScioli et al. 2011b), an important factor in being able to predict others' actions in the kind of triadic conflicts Pietraszewski has in mind.

Indeed, the problem is even worse than that because humans do not always take the side of close allies when conflicts emerge, instead using the moral value of the acts of those involved in the conflict to choose sides (DeScioli 2008; DeScioli et al. 2011a; Kurzban et al. 2012). So, although we agree with **Uhlmann** that often people desire that those who violate a social norm should be punished even if the target or the target's allies were not harmed (Kurzban et al. 2007), we believe – and suspect that **Pahlavan** would agree – that it will be useful to distinguish revenge from moralistic aggression, which we believe evolved for a different function (Szymanska 2011).

We are similarly skeptical of **Uhlmann**'s suggestion that people's reactions to harmless moral violations are designed around inferences about character (DeScioli et al. 2012; Gray et al., in press) – why the accompanying desire for punishment as opposed to mere avoidance, if the goal is simply character evaluation? – but we agree that this remains an open and interesting question (DeScioli & Kurzban 2009a).

For these reasons, we are optimistic that the overall approach we suggest here might be fruitfully applied to groups, but we continue to believe that additional theoretical development will be needed to bridge the gap from dyad to collective.

R8. What are the implications for criminal justice and restorative justice?

We appreciate **Roberts & Murray**'s, and **Petersen**'s suggestions that our ideas about revenge might be productively applied to understanding how humans make decisions in the context of contemporary criminal justice systems. Roberts & Murray note that when people empathize with victims (e.g., when the costs to the victim are high), they are more likely to "view the costs of the crime in a personal manner," leading to an increase in their motivation to retaliate on behalf of victims via their decisions about guilt versus innocence or about punishment severity. We appreciate this point, and Petersen shows that even more subtle predictions are possible. In particular, he notes that third parties will inevitably disagree about how much punishment versus restorative action a specific offender should receive because social value is a dyadic phenomenon. As Petersen writes, "different selves will compute the social value of the same target differently, and a single self will compute the social value of different targets differently." Consequently, disagreement among jurors or other third parties regarding the sanctions that should be delivered to a criminal (as a function, say, of racial or ethnic similarity, or other cues of social value), particularly in information-sparse decision-making contexts where the ancestrally valid cues for making such determinations are generally unavailable, is not only unsurprising, but probably also inevitable. The applications of this insight, both for research and policy, are considerable.

R9. Do people ever genuinely forgive?

Finally, **McCoy & Shackelford** raise doubts as to whether people ever "genuinely" forgive. We actually agree with their analysis suggesting that systems for regulating behavior with respect to others are unlikely to be designed to discard entirely past information about harms; indeed, past information should be used in some more or less Bayesian fashion to update expectations. Our proposal regarding forgiveness was not meant to imply that the past behavior did not influence subsequent behavior in any way. In fact, we take for granted that a "genuine forgiver" will also be "prepared to exact revenge" if a future harm occurs in the same way that McCoy & Shackelford envision their so-called "feigned forgiver" would be. Our model does not imply that someone who forgives at time one necessarily will permanently forego revenge in the future. Moreover, we note, after Petersen et al. (2010) that forgiveness-based strategies for addressing exploitation are dicey propositions because failures to impose retaliatory harms can easily be confused for weakness or failures of nerve, inviting further exploitation from one's harmdoer.

Still, we doubt that all forgiveness signals are disingenuous (i.e., serving only to lull others into complacency) for two principal reasons. First, if all such signals were false, then they would be ignored, for the usual reasons associated with the evolution of signaling systems (Dawkins & Krebs 1978). Second, we believe that harms occur within the context of relationships that are, with some probability, likely to produce positive sum outcomes over time, which means that reconciliation along the lines we propose carries a higher expected value in some contexts than exacting revenge would. We think that forgiveness might serve to *put relationships (back) on a positive sum footing while still leaving open the possibility of revenge, should further harms occur.*

R10. Conclusion

Once again, we wish to thank the many commentators who responded so thoughtfully to our target article. We see substantial (though hardly unanimous) agreement among our colleagues about the basic framework we propose – specifically, that revenge is a deterrence system and forgiveness is a system designed to preserve valuable relationships.

This is not, of course, to minimize the work still left to be done. Empirical research that carefully evaluates the computational systems responsible for generating revenge and forgiveness would be most welcome, as would continuing work on the development of these systems over the life course and the interaction of individual (e.g., sex, personality, genetic), cultural, and ecological differences with the computational systems we have sketched here (McCullough 2008). In addition, as imaging technology becomes more powerful and theorizing about the interface of cognitive science and neuroscience becomes more sophisticated, cognitive neuroscientists will increasingly be in a position to shed light on the neural bases of the computational systems we have presented here. Finally, we look forward to the possibility that the ideas we have presented might help to build bridges to the work of our colleagues in biology who study the evolution and operation of homologous behavioral systems in nonhuman species.

References

[The letters "a" and "r" before author's initials stand for target article and response references, respectively]

Abwender, D. A. & Hough, K. (2001) Interactive effects of characteristics of defendant and mock juror on US participants' judgment and sentencing recommendations. *Journal of Social Psychology* 141:603–15. [SCR]

Adinkrah, M. (2011) Child witch hunts in contemporary Ghana. *Child Abuse and Neglect* 35:741–52. [MP]

Alicke, M. D. (1992) Culpable causation. *Journal of Personality and Social Psychology* 63:368–78. [KO]

Allan, A., Allan, M. M., Kaminer, D. & Stein, D. J. (2006) Exploration of the association between apology and forgiveness amongst victims of human rights violations. *Behavioral Science and the Law* 24(1):87–102. [aMEMcC, DJS]

Allport, G. W. (1954) *The nature of prejudice*. Doubleday. [RJC]

Anderson, C. & Putterman, L. (2006) Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior* 54:1–24. [aMEMcC]

Anderson, E. (1999) *Code of the street*. W. W. Norton. [aMEMcC]

Anderson, M. L. (2010) Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences* 33:245–313. [rMEMcC]

Andreoni, J., Harbaugh, W. & Vesterlund, L. (2003) The carrot or the stick: Rewards, punishments, and cooperation. *American Economic Review* 93:893–902. [aMEMcC]

Andrews, P. W., Gangestad, S. W. & Matthews, D. (2002) Adaptationism – How to carry out an exaptationist program. *Behavioral and Brain Sciences* 25:489–504. [arMEMcC]

Archer, J. (2004) Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology* 8:291–322. [aMEMcC]

Archer, J. (2009) Does sexual selection explain human sex differences in aggression? *Behavioral and Brain Sciences* 32:249–311. [aMEMcC]

Asendorpf, J. B., Banse, R. & Mücke, D. (2002) Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology* 83(2):380–93. [RF]

Asmal, K., Asmal, L. & Roberts S, R. (1996) *Reconciliation through truth: A reckoning of apartheid's criminal governance*. David Phillip. [DJS]

Aureli, F. (1997) Post-conflict anxiety in nonhuman primates: The mediating role of emotion in conflict resolution. *Aggressive Behavior* 23:315–28. [aMEMcC]

Aureli, F., Cozzolino, R., Cordischi, C. & Scucchi, S. (1992) Kin-oriented redirection among Japanese macaques: An expression of a revenge system? *Animal Behaviour* 44:283–91. [FA, arMEMcC]

Aureli, F., Fraser, O. N., Schaffner, C. M. & Schino, G. (in press) The regulation of social relationships. In: *Evolution of primate societies*, ed. J. Mitani, J. Call, P. Kappeler, R. Palombit & J. Silk. University of Chicago Press. [FA]

Aureli, F. & de Waal, F. B. M. (2000) *Natural conflict resolution*. University of California Press. [aMEMcC]

Aureli, F. & Schaffner, C. M. (2002) Relationship assessment through emotional mediation. *Behaviour* 139:393–420. [FA, aMEMcC]

Aureli, F. & Schino, G. (2004) The role of emotions in social relationships. In: *Macaque societies: A model for the study of social organization*, ed. B. Thierry, M. Singh & W. Kaumanns, pp. 38–55. Cambridge University Press. [FA]

Aureli, F. & van Schaik, C. P. (1991a) Post-conflict behaviour in long-tailed macaques (*Macaca fascicularis*): I. The social events. *Ethology* 89(2):89–100. [AD]

Aureli, F. & van Schaik, C. P. (1991b) Post-conflict behaviour in long-tailed macaques (*Macaca fascicularis*): II. Coping with the uncertainty. *Ethology* 89 (2):101–14. [aMEMcC]

Aureli, F. & Whiten, A. (2003) Emotions and behavioral flexibility. In: *Primate psychology – The mind and behavior of human and nonhuman primates*, ed. D. Maestripieri, pp. 289–323. Harvard University Press. [FA]

Axelrod, R. (1980) More effective choice in the Prisoner's Dilemma. *Journal of Conflict Resolution* 24:379–403. [aMEMcC]

Axelrod, R. (1984) *The evolution of cooperation*. Basic Books. [PB, aMEMcC]

Ballard, K. & Knutson, B. (2009) Dissociable neural representations of future reward magnitude and delay during temporal discounting. *NeuroImage* 45:143–50. [rMEMcC]

Balliet, D. (2010) Conscientiousness and forgivingness: A meta-analysis. *Personality and Individual Differences* 48:259–63. [DB, rMEMcC]

Balliet, D. & Joireman, J. (2010) Ego depletion reduces proselfs' concern with the well-being of others. *Group Processes and Intergroup Relations* 13:227–39. [DB]

Balliet, D., Li, N. P. & Joireman, J. (2011a) Relating trait self-control and forgiveness among prosocials and proselfs: A test of compensatory and synergistic models. *Journal of Personality and Social Psychology* 101:1090–109. [DB]

Balliet, D., Mulder, L. B. & Van Lange, P.A.M. (2011b) Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin* 137:594–615. [DB]

Balliet, D., Parks, C. & Joireman, J. (2009) Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes and Intergroup Relations* 12:533–47. [DB]

Bandura, A. (1999) Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review* 3:193–209. [FP]

Barclay, P. (2006) Reputational benefits for altruistic punishment. *Evolution and Human Behavior* 27:325–44. [aMEMcC]

Barclay, P. (2008) Using the hatchet and burying it afterwards – A review of *Beyond revenge: The evolution of the forgiveness instinct*. *Evolution and Human Behavior* 29:450–51. [PB]

Barclay, P. (2011) The evolution of charitable behaviour and the power of reputation. In: *Applied evolutionary psychology*, ed. C. Roberts. pp. 149–72. Oxford University Press. [PB]

Barnes, C. D., Brown, R. P. & Osterman, L. L. (2009) Protection, payback, or both? Emotional and motivational mechanisms underlying avoidance by victims of transgressions. *Motivation and Emotion* 33:400–11. [aMEMcC]

Baron-Cohen, S. (1995) *Mindblindness: An essay on autism and theory of mind.* Cambridge University Press. [TJW]

Barrett, H. C. & Kurzban, R. (2006) Modularity in cognition: Framing the debate. *Psychological Review* 113:628–47. [rMEMcC]

Barrett, L., Gaynor, D. & Henzi, P. (2002) A dynamic interaction between aggression and grooming reciprocity among female chacma baboons. *Animal Behaviour* 63:1047–53. [AD]

Baskin, T. W. & Enright, R. D. (2004) Intervention studies on forgiveness: A meta-analysis. *Journal of Counseling and Development* 82:79–90. [aMEMcC]

Batson, C. D. (2011) *Altruism in humans*. Oxford. [aMEMcC]

Batson, C. D., Kennedy, C. L., Nord, L., Stocks, E. L., Fleming, D. A., Marzette, C. M., Lishner, D. A., Hayes, R. E., Kolchinsky, L. M., & Zerger, T. (2007) Anger at unfairness: Is it moral outrage? *European Journal of Social Psychology* 37:1272–85. [aMEMcC]

Baumeister, R. F., Stillwell, A. & Wotman, S. R. (1990) Victim and perpetrator accounts of interpersonal conflict: Autobiographical narratives about anger. *Journal of Personality and Social Psychology* 59(5):994–1005. [SK]

Bayer, C. P., Klasen, F. & Adam, H. (2007) Association of trauma and PTSD symptoms with openness to reconciliation and feelings of revenge among former Ugandan and Congolese child soldiers. *Journal of the American Medical Association* 298:555–59. [aMEMcC]

Becker, K. D., Steuwig, J., Herrera, V. M. & McCloskey, L. A. (2004) A study of firesetting and animal cruelty in children: Family influences and adolescent outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry* 43:905–12. [ELU]

Beckerman, S., Erickson P. I., Yost, J., Regalado, J., Jaramillo, L., Sparks, C., Iromenga, M. & Long, K. (2009) Life histories, blood revenge, and reproductive success among the Waorani of Ecuador. *Proceedings of the National Academy Sciences USA* 106:8134–39. [aMEMcC, MP]

Belsky, J., Steinberg, L. & Draper, P. (1991) Childhood experience, interpersonal development, and reproductive strategy: An evolutionary theory of socialization. *Child Development* 62:647–70. [rMEMcC]

Bendor, J., Kramer, R. M. & Stout, S. (1991) When in doubt: Cooperation in a noisy prisoner's dilemma. *Journal of Conflict Resolution* 35:691–719. [aMEMcC]

Benenson, J. F., Carder, H. P. & Geib-Cole, S. J. (2008) The development of boys' preferential pleasure in physical aggression. *Aggressive Behavior* 34:154–66. [MP]

Bentham, J. (1962) Principles of penal law. In: *The works of Jeremy Bentham*, ed. J. Bowring, pp. 365–580. Russell and Russell. [aMEMcC]

Berg, J., Dickhaut, J. & McCabe, K. (1995) Trust, reciprocity and social history. *Games and Economic Behavior* 10:122–42. [aMEMcC]

Bernhard, H., Fischbacher, U. & Fehr, E. (2006) Parochial altruism in humans. *Nature* 442:912–15. [aMEMcC]

Bettencourt, B. A. & Miller, N. (1996) Gender differences in aggression as a function of provocation: A meta-analysis. *Psychological Bulletin* 119:422–47. [aMEMcC]

Bies, R. J. & Tripp, T. M. (1996) Beyond distrust: "Getting even" and the need for revenge. In: *Trust in organizations: Frontiers of theory and research,* ed. M. R. Kramer & T. R. Tyler, pp. 246–60. Sage. [KO]

Bixenstine, V. E. & Wilson, K. V. (1963) Effects of level of cooperative choice by the other player on choices in a prisoner's dilemma game. Part II. *Journal of Abnormal and Social Psychology* 67:139–47. [aMEMcC]

Black-Michaud, J. (1975) *Cohesive force: Feud in the Mediterranean and the Middle East*. Basil Blackwell. [aMEMcC]

Blumstein, D. T., Ferando, E. & Stankowich, T. (2009) A test of the multipredator hypothesis: Yellow-bellied marmots respond fearfully to the sight of novel and extinct predators. *Animal Behaviour* 78:873–78. [rMEMcC]

Bochet, O., Page, T. & Putterman, L. (2006) Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior and Organization* 60:11–26. [aMEMcC]

Boehm, C. (1987) *Blood revenge: The enactment and management of conflict in Montenegro and other tribal societies*, 2nd edition. University of Pennsylvania Press. [aMEMcC]

Boehm, C. (2008) Purposive social selection and the evolution of human altruism. *Cross-Cultural Research* 42:319–52. [aMEMcC]

Boehm, C. (2011) *Moral origins: The evolution of virtue, altruism, and shame.* Basic Books. [HG]

Boesch, C. (2007) What makes us human (Homo sapiens)? The challenge of cognitive cross-species comparison. *Journal of Comparative Psychology* 121:227–40. [TJW]

Bohnet, I., Herrmann, B. & Zeckhauser, R. (2010) Trust and the reference points for trustworthiness in Gulf and Western countries. *The Quarterly Journal of Economics* 125(2):811–28. [DL]

Bohnet, I. & Zeckhauser, R. (2004) Trust, risk and betrayal. *Journal of Economic Behavior and Organization* 55:467–84. [DL]

Bolton, G. E. & Zwick, R. (1995) Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior* 10:95–121. [aMEMcC]

Bornstein, G. (2003) Intergroup conflict: Individual, group, and collective interests. *Personality and Social Psychology Review* 7:129–45. [RJC]

Boster, J. S., Yost, J. & Catherine Peeke, C. (2004) Rage, revenge, and religion: Honest signaling of aggression and nonaggression in Waorani coalitional violence. *Ethos* 31:471–94. [MP]

Bottom, W. P., Gibson, K., Daniels, S. E. & Murnighan, J. K. (2002) When talk is not cheap: Substantive penance and expressions of intent in rebuilding cooperation. *Organization Science* 13:497–513. [aMEMcC]

Boulding, K. E. (1969) Economics as a moral science. *The American Economic Review* 59(1):1–12. [DJS]

Bowles, S. & Gintis, H. (2011) *A cooperative species: Human reciprocity and its evolution.* Princeton University Press. [HG]

Boyd, R., Gintis, H. & Bowles, S. (2010) Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* 328:617–20. [HG]

Bramel, D., Taub, B. & Blum, B. (1968) An observer's reaction to the suffering of his enemy. *Journal of Personality and Social Psychology* 8:384–92. [VJK]

Brezina, T., Agnew, R., Cullen, F. T. & Wright, J. P. (2004) The code of the street: A quantitative assessment of Elijah Anderson's subculture of violence thesis and its contribution to youth violence research. *Youth Violence and Juvenile Justice* 2:303–28. [aMEMcC]

Brickman, P., Rabinowitz, V. C., Karuza, J., Coates, D., Cohen, E. & Kidder, L. (1982) Models of helping and coping. *American Psychologist* 37:368–84. [FP]

Brown, B. R. (1968) The effects of need to maintain face on interpersonal bargaining. *Journal of Experimental Social Psychology* 4:107–22. [aMEMcC]

Brown, D. E. (1991) *Human universals*. McGraw-Hill. [aMEMcC]

Brown, R. & Hewstone, M. (2005) An integrative theory of intergroup contact. In: *Advances in experimental social psychology,* ed. M. Zanna, pp. 255–343. Academic Press. [RJC]

Bshary, A. & Bshary, R. (2010) Self-serving punishment of a common enemy creates a public good in reef fishes. *Current Biology* 20:1–4. [rMEMcC]

Bshary, R. & Grutter, A. S. (2005) Punishment and partner switching cause cooperative behavior in a cleaning mutualism. *Biology Letters* 1:396–99. [arMEMcC]

Buchanan, T. W. (2007) Retrieval of emotional memories. *Psychological Bulletin* 133 (5):761–79. doi: 10.1037/0033-2909.133.5.761 [RF]

Buchholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D. & Marois, R. (2008) The neural correlates of third- party punishment. *Neuron* 930–40. [HG]

Buford, B. (1992) *Among the thugs: The experience, and seduction, of crowd violence.* W. W. Norton. (Original Vintage edition published in 1990). [MP]

Bureau of Alcohol, Tobacco, and Firearms (1999) *Arson and Explosives Incidents Report, 1997.* (No. ATF P 3320.4). [aMEMcC]

Burnette, J. L., McCullough, M. E., Van Tongeren, D. R. & Davis, D. E. (2012) Forgiveness results from integrating information about relationship value and exploitation risk. *Personality and Social Psychology Bulletin* 38:345–56. [rMEMcC, MBP]

Burnham, T. C. (2003) Engineering altruism: A theoretical and experimental investigation of anonymity and gift giving. *Journal of Economic Behavior and Organization* 50:133–44. [arMEMcC]

Burnham, T. C. & Johnson, D. D. P. (2005) The biological and evolutionary logic of human cooperation. *Analyse & Kritik* 27:113–35. [aMEMcC]

Buss, D. M. (2009) How can evolutionary psychology successfully explain personality and individual differences? *Perspectives on Psychological Science* 4:359–66. [arMEMcC]

Call, J., Aureli, F. & de Waal, F. B. M. (1999) Reconciliation patterns among stumptailed macaques: A multivariate approach. *Animal Behaviour* 58:165–72. [aMEMcC]

Camerer, C. F. (2003) *Behavioral game theory*. Princeton University Press. [aMEMcC]

Campbell, W. K. & Foster, C. A. (2002) Narcissism and commitment in romantic relationships: An investment model analysis. *Personality and Social Psychology Bulletin* 28(4):484–95. [SK]

Campbell, W. K., Foster, C. & Finkel, E. J. (2002) Does self-love lead to love for others? A story of narcissistic game playing. *Journal of Personality and Social Psychology* 83(2):340–54. [SK]

Cant, M. A. (2011) The role of threats in animal cooperation. *Proceedings of the Royal Society of London B: Biological Sciences* 278:170–78. [aMEMcC]
Cant, M. A. & Johnstone, R. A. (2006) Self-serving punishment and the evolution of cooperation. *Journal of Evolutionary Biology* 19:1383–85. [aMEMcC]
Caporael, L. R. (1997) The evolution of truly social cognition: The core configurations model. *Personality and Social Psychology Review* 1:276–98. [RJC]
Caporael, L. R., Baron, R. M. (1997) Groups as the minds natural environment. In: *Evolutionary social psychology,* ed. J. A. Simpson & D. T. Kenrick, pp. 317–45. Erlbaum. [RJC]
Carcach, C. (1997) *Youth as victims and offenders of homicide*, No. 73. Australian Institute of Criminology. [aMEMcC]
Cardona, M., Garcia, H. I., Giraldo, C. A., Lopez, M. V., Suarez, C. M., Corcho, D. C., Posada, C. H., & Florez, M. N. (2005) Homicides in Medellin, Colombia, from 1990 to 2002: Victims, motives, and circumstances. *Cadernos de Salude Publica, Rio de Janeiro* 21:840–51. [aMEMcC]
Carlsmith, K. M. & Darley, J. M. (2008) Psychological aspects of retributive justice. In: *Advances in experimental social psychology,* ed. M. P. Zanna, 41:193–36. [KO]
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* 83:284–99. [KO]
Carlsmith, K. M., Wilson, T. D. & Gilbert, D. T. (2008) The paradoxical consequences of revenge. *Journal of Personality and Social Psychology* 95:1316–24. [aMEMcC, FP]
Carlson, M. & Miller, N. (1988) Bad experiences and aggression. *Sociology and Social Research* 72:155–58. [aMEMcC]
Carmichael, C. M. & McGue, M. (1995) A cross-sectional examination of height, weight, and body mass index in adult twins. *Journal of Gerontology Series A: Biological and Medical Sciences* 50:B237–44. [aMEMcC]
Carpendale, J. I. M. & Lewis, C. (2006) *How children develop social understanding.* Blackwell. [TJW]
Carpenter, J. P. & Matthews, P. H. (2004) Social reciprocity. Middlebury College Department of Economics, working paper. [aMEMcC]
Carpenter, J. P. & Matthews, P. H. (2009) What norms trigger punishment? *Experimental Economics* 12:272–88. [aMEMcC]
Carpenter, J. P. & Matthews, P. H. (2012) Norm enforcement: Anger, indignation, or reciprocity? *Journal of the European Economic Association*. 10:555–72. [aMEMcC]
Carruthers, P. (2006) *The architecture of the mind*. Clarendon Press. [rMEMcC]
Castles, D. L. & Whiten, A. (1998a) Post-conflict behaviour of Wild Olive Baboons. I. Reconciliation, redirection and consolation. *Ethology* 104(2):126–47. [AD]
Castles, D. L. & Whiten, A. (1998b) Post-conflict behaviour of Wild Olive baboons. II. Stress and self-directed behaviour. *Ethology* 104:148–60. [aMEMcC]
Chagnon, N. (1988) Life histories, blood revenge, and warfare in a tribal population. *Science,* 239:985–92. [aMEMcC]
Charney, E. (2012) Behavior genetics and postgenomics. *Behavioral and Brain Sciences*. 35(5):331–410. [TJW]
Cheney, D. L. & Seyfarth, R. M. (2007) *Baboon metaphysics: The evolution of a social mind*. University of Chicago Press. [aMEMcC]
Cherniak, C., Mokhtarzada, Z., Rodriguez-Esteban, R. & Changizi, K. (2004) Global optimization of cerebral cortex layout. *Proceedings of the National Academy of Sciences USA* 101:1081–86. [rMEMcC]
Cinyabuguma, M., Page, T. & Putterman, L. (2006) Can second-order punishment deter perverse punishment? *Experimental Economics* 9:265–79. [aMEMcC]
Clark, K. & Sefton, M. (2001) The sequential prisoner's dilemma: Evidence on reciprocation. *The Economic Journal* 111:51–68. [aMEMcC]
Clavien, C. & Klein, R. A. (2010) Eager for fairness or for revenge? Psychological altruism in economics. *Economics and Philosophy* 26:267–90. [aMEMcC]
Clutton-Brock, T. H. (1989) Mammalian mating systems. *Proceedings of the Royal Society B:* 236:339–72. [aMEMcC]
Clutton-Brock, T. H. & Parker, G. A. (1995) Punishment in animal societies. *Nature* 373:209–16. [FA, aMEMcC]
Cohen, D., Nisbett, R. E., Bowdle, B. F. & Schwarz, N. (1996) Insult, aggression, and the Southern culture of honor: An "experimental ethnography." *Journal of Personality and Social Psychology* 70:945–60. [aMEMcC]
Cools, A. K. A., Van Hout, A. J.-M. & Nelissen, M. H. J. (2008) Canine reconciliation and third-party-initiated postconflict affiliation: Do peacemaking social mechanisms in dogs rival those of higher primates? *Ethology* 114:53–63. [aMEMcC]
Cooper, M. A., Bernstein, I. S. & Hemelrijk, C. K. (2005) Reconciliation and relationship quality in Assamese macaques (*Macaca assamensis*). *American Journal of Primatology* 65:269–82. [aMEMcC]
Copeland-Linder, N., Jones, V. C., Haynie, D. L., Simons-Morton, B. G., Wright, J. L. & Cheng, T. L. (2007) Factors associated with retaliatory attitudes among African American adolescents who have been assaulted. *Journal of Pediatric Psychology* 32:760–70. [aMEMcC]
Cordoni, G. & Palagi, E. (2008) Reconciliation in Wolves (*Canis lupus*): New evidence for a comparative perspective. *Ethology* 114:298–308. [aMEMcC]
Correia, I. & Claudia Dalbert, C. (2008) School bullying: Belief in a personal just world of bullies, victims, and defenders. *European Psychologist* 13(4):248–54. DOI: 10.1027/1016-9040.13.4.248 [FP]
Coser, L. A. (1956) *The function of social conflict*. Free Press. [RJC]
Courtois, G., ed. (1984) *La Vengeance: Études d'ethnologies, d'histoire et de philosophie, vol. 4: La vengeance dans la pensée occidentale.* Éditions Cujas. [SB]
Crisp, R. J. & Hewstone, M. (2007) Multiple social categorization. In: *Advances in experimental social psychology,* ed. M. Zanna, pp. 163–254. Academic Press. [RJC]
Crisp, R. J. & Turner, R. N. (2009) Can imagined interactions produce positive perceptions? Reducing prejudice through simulated social contact. *American Psychologist* 64:231–40. [RJC]
Crisp, R. J. & Turner, R. N. (2011) Cognitive adaptation to the experience of social and cultural diversity. *Psychological Bulletin* 137:242–66. [RJC]
Crook, J. H. (1989) Introduction – Socioecological paradigms, evolution and history: Perspectives for the 1990s. In: *Comparative socioecology: The behavioural ecology of humans and other mammals*, ed. V. Standen & R. Foley, pp. 1–36. Blackwell Scientific. [FA]
Cullen, F. T., Fisher, B. S. & Applegate, B. K. (2000) Public opinion and crime and corrections. *Crime and Justice* 27:1–79. [MBP]
Cullen, F. T. & Gendreau, P. (2000) Assessing correctional rehabilitation: Policy, practice, and prospects. *Criminal Justice* 3:109–75. [MBP]
Curry, O. S., Price, M. E. & Price, J. G. (2008) Patience is a virtue: Cooperative people have lower discount rates. *Personality and Individual Differences* 44:780–85. [rMEMcC]
Cushman, F. (2008) Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 108:353–80. [KO]
Daly, M. & Wilson, M. (1988) *Homicide*. Aldine de Gruyter. [PB, aMEMcC]
Daly, M. & Wilson, M. (2005) Carpe diem: Adaptation and devaluing the future. *Quarterly Review of Biology* 80:55–60. [rMEMcC]
Damasio, A. R. (1994) *Descartes' error – Emotion, reason, and the human brain*. Grosset/Putnam. [FA]
Damasio, A. R. (1996) The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London B* 351:1413–20. [FA]
Daniel, J. R., Santos, A. J. & Cruz, M. G. (2009) Postconflict behaviour in brown capuchin monkeys (*Cebus apella*). *Folia Primatologica* 80:329–40. [aMEMcC]
Darley, J. M. & Pittman, T. S. (2003) The psychology of compensatory and retributive justice. *Personality and Social Psychology Review* 7:324–36. [KO]
Darley, J. M., Carlsmith, K. M. & Robinson, P. H. (2000) Incapacitation and just deserts as motives for punishment. *Law and Human Behavior* 24:659–83. [KO]
Davis, D. E., Worthington, E. L., Jr., Hook, J. N., Van Tongeren, D. L., Green, J. D. & Jennings, D. J. (2009) Relational spirituality and the development of the Similarity of the Offender's Spirituality Scale. *Psychology of Religion and Spirituality* 1:249–62. [aMEMcC]
Dawes, R. M. (1980) Social dilemmas. *Annual Review of Psychology* 31:169–93. [RJC]
Dawkins, R. & Krebs, J. R. (1978) Animal signals: Information or manipulation? In: *Behavioural ecology: An evolutionary approach,* 1st edition, ed. J. R. Krebs & N. B. Davies, pp. 282–309. Blackwell. [rMEMcC]
De Cremer, D. & Van Lange, P. A. M. (2001) Why prosocials exhibit greater cooperation than proselfs: The roles of social responsibility and reciprocity. *European Journal of Personality* 15(Suppl. 1):S5–S18. [DL]
de Jong, P. J., Peters, M. L. & de Cremer, D. (2003) Blushing may signify guilt: Revealing effects of blushing in ambiguous social situations. *Motivation and Emotion* 27:225–49. [aMEMcC]
Delton, A. W., Krasnow, M. M., Cosmides, L. & Tooby, J. (2011) Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences USA* 108:13335–40. [rMEMcC]
Denant-Boemont, L., Masclet, D. & Noussair, C. N. (2007) Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory* 33:145–67. [aMEMcC]
DePrince, A. P., Brown, L. S., Cheit, R. E., Freyd, J. J., Gold, S. N., Pezdek, K. & Quina, K. (2012) Motivated forgetting and misremembering: Perspectives from Betrayal Trauma Theory. In: *True and false recovered memories: Toward a reconciliation of the debate. Nebraska Symposium on Motivation 58*, ed. R. F. Belli, pp. 193–243. Springer. [SJ-F]
de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. & Fehr, E. (2004) The neural basis of altruistic punishment. *Science* 305(5688):1254–58. [DJS]
DeScioli, P. (2008) *Investigations into the problems of moral cognition.* Doctoral Dissertation, University of Pennsylvania. [rMEMcC]
DeScioli, P., Bruening, R. & Kurzban, R. (2011a) The omission effect in moral cognition: Toward a functional explanation. *Evolution and Human Behavior* 32:204–15. [rMEMcC]

DeScioli, P. & Kurzban, R. (2009a) Mysteries of morality. *Cognition* 112:281–99. [rMEMcC]

DeScioli, P. & Kurzban, R. (2009b) The alliance hypothesis for human friendship. *PLoS One* 4:e5802. [arMEMcC]

DeScioli, P. & Kurzban, R. (2012) The company you keep: Friendship decisions from a functional perspective. In: *Social judgment and decision making*, ed. J. I. Krueger. Psychology Press. [rMEMcC]

DeScioli, P., Gilbert, S. & Kurzban, R. (2012) Indelible victims and persistent punishers in moral cognition. *Psychological Inquiry* 23:143–49. [rMEMcC]

DeScioli, P., Kurzban, R., Koch, E. N. & Liben-Nowell, D. (2011b) Best friends: Alliances, friend ranking, and the MySpace social network. *Perspectives on Psychological Science* 6:6–8. [rMEMcC]

Deutsch, M. (1973) *The resolution of conflict*. Yale University Press. [RJC]

de Waal, F. B. M. (1997) *Good natured: The origins of right and wrong in humans and other animals.* Harvard University Press. [HG]

de Waal, F. B. M. (2000) Primates – A natural heritage of conflict resolution. *Science* 289(5479):586–90. [aMEMcC, DJS]

de Waal, F. B. M. & van Roosmalen, A. (1979) Reconciliation and consolation among chimpanzees. *Behavioral Ecology and Sociobiology* 5:55–66. [aMEMcC]

DeWall, C. N., Baumeister, R. F., Stillman, T. F. & Gailliot, M. T. (2007) Violence restrained: Effects of self-regulation and its depletion on aggression. *Journal of Experimental Social Psychology* 43:62–76. [aMEMcC, RY]

DeWall, C. N., Pond, R. S., Jr. & Bushman, B. J. (2010) Sweet revenge: Diabetic symptoms predict less forgiveness. *Personality and Individual Differences* 49:823–26. [RY]

Diamond, S. R. (1977) The effect of fear on the aggressive responses of anger aroused and revenge motivated subjects. *Journal of Psychology* 95:185–88. [aMEMcC]

DiBlasio, F. & Benda, B. B. (1991) Practitioners, religion and the use of forgiveness in the clinical setting. *Journal of Psychology and Christianity* 10:166–72. [aMEMcC]

Dijk, C., de Jong, P. J. & Peters, M. L. (2009) The remedial value of blushing in the context of transgressions and mishaps. *Emotion* 9:287–91. [aMEMcC]

Doob, A. N. & Wood, L. E. (1972) Catharsis and aggression: Effects of annoyance and retaliation on aggressive behavior. *Journal of Personality and Social Psychology* 22:156–62. [VJK]

Dooley, E. (2001) *Homicide in Ireland 1992–1996*. Government Report No. 9435. Government of Ireland. [aMEMcC]

dos Santos, M., Rankin, D. J. & Wedekind, C. (2011) The evolution of punishment through reputation. *Proceedings of the Royal Society of London, B* 278:371–77. [aMEMcC]

Dreber, A., Rand, D. G., Fudenberg, D. & Nowak, M. A. (2008) Winners don't punish. *Nature* 452(7185):348–51. [aMEMcC, RY]

Dunbar, R. I. M., Clark, A. & Hurst, N. L. (1995) Conflict and cooperation among the Vikings: Contingent behavioral decisions. *Ethology and Sociobiology* 16:233–46. [aMEMcC]

Durkheim, E. (1902/1967) *De La division du travail social.* Presses Universitaires de France. [HG]

Eaton, J. & Struthers, C. W. (2006) The reduction of psychological aggression across varied interpersonal contexts through repentance and forgiveness. *Aggressive Behavior* 32:195–206. [aMEMcC]

Eaton, J., Struthers, C. W. & Santelli, A. G. (2006) The mediating role of perceptual validation in the repentance-forgiveness process. *Personality and Social Psychology Bulletin* 32:1389–401. [aMEMcC]

Eaves, L. J., Hatemi, P. K., Prom-Womley, E. C. & Murrelle, L. (2008) Social and genetic influences on adolescent religious attitudes and practices. *Social Forces* 86:1621–46. [aMEMcC]

Egas, M. & Riedl, A. (2008) The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B* 275:871–78. [aMEMcC]

Eid, M. & Diener, E. (2006) Introduction: The need for multimethod measurement in psychology. In: *Handbook of multimethod measurement in psychology*, ed. M. Eid & E. Diener, pp. 3–8. American Psychological Association. [RF]

Eisenberg, N. & Lennon, R. (1983) Sex differences in empathy and related capacities. *Psychological Bulletin* 94(1):100–31. [SK]

Eisenberger, R., Lynch, P., Aselage, J. & Rohdieck, S. (2004) Who takes the most revenge? Individual differences in negative reciprocity norm endorsement. *Personality and Social Psychology Bulletin* 30:787–99. [aMEMcC]

Ekman, P. & O'Sullivan, M. (2006) From flawed self-assessment to blatant whoppers: The utility of voluntary and involuntary behavior in detecting deception. *Behavioral Sciences and the Law* 24:673–86. [SJ-F]

Ellis, B. J. (2004) Timing of pubertal maturation in girls: An integrated life history approach. *Psychological Bulletin* 130:920–58. [rMEMcC]

Ellis, B. J. & Bjorklund, D. F. eds. (2005) *Origins of the social mind: Evolutionary psychology and child development*. Guilford Press. [rMEMcC]

Ellsworth, P. C. & Ross, L. D. (1983) Public opinion and Capital Punishment: A close examination of the views of Abolitionists and Retentionists. *Crime and Delinquency* 29:116–69. [KO]

Ellsworth, P. C. & Scherer, K. R. (2003) Appraisal processes in emotion. In: *Handbook of affective sciences,* ed. R. J. Davidson, H. Goldsmith & K. R. Scherer, pp. 572–95. Oxford University Press. [DL]

Elster, J. (1990) Norms of revenge. *Ethics* 100:862–85. [aMEMcC]

Engel, C. (in press) Dictator games: A meta-study. *Experimental Economics*. [aMEMcC]

Engh, A. L., Siebert, E. R., Greenberg, D. A. & Holekamp, K. E. (2005) Patterns of alliance formation and postconflict aggression indicate spotted hyaenas recognize third-party relationships. *Animal Behaviour* 69:209–17. [aMEMcC]

Enright, R. D. & Coyle, C. T. (1998) Researching the process model of forgiveness within psychological interventions. In: *Dimensions of forgiveness: Psychological research and theological perspectives,* ed. E. L. Worthington, Jr., pp. 139–61. Templeton Foundation Press. [aMEMcC]

Ericksen, K. P. & Horton, H. (1992) "Blood feuds": Cross-cultural variations in kin group vengeance. *Behavior Science Research* 26:57–85. [aMEMcC]

Evans-Pritchard, E. E. (1940) *The nuer*. Oxford University Press. [RJC]

Exline, J. J. & Baumeister, R. F. (2000) Expressing forgiveness and repentance: Benefits and barriers. In: *Forgiveness: Theory, research, and practice*, ed. M. E. McCullough, K. I. Pargament & C. E. Thoresen, pp. 133–55. Guilford Press. [RF]

Exline, J. J., Baumeister, R. F., Bushman, B. J., Campbell, W. K. & Finkel, E. J. (2004) Too proud to let go: Narcissistic entitlement as a barrier to forgiveness. *Journal of Personality and Social Psychology* 87(6):894–912. [SK]

Farrow, T. F., Hunter, M. D., Wilkinson, I. D., Gouneea, C., Fawbert, D., Smith, R., Lee, K. H., Mason, S., Spence, S. A. & Woodruff, P. W. (2005) Quantifiable change in functional brain response to empathic and forgivability judgments with resolution of posttraumatic stress disorder. *Psychiatry Research* 140(1):45–53. [RF]

Farrow, T. F., Zheng, Y., Wilkinson, I. D., Spence, S. A., Deakin, J. F., Tarrier, N., Griffiths, P. D. & Woodruff, P. W. (2001) Investigating the functional anatomy of empathy and forgiveness. *NeuroReport* 12(11):2433–38. [RF]

Federici, S. (2010) Women, witch-hunting and enclosures in Africa today. *Sozial. Geschichte Online* 3(Suppl.):10–27. Available at: http://www.stiftung-sozialgeschichte.de [MP]

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

Fehr, E. & Gachter, S. (1998) How effective are trust- and reciprocity-based incentives? In: *Economics, values and organizations*, ed. Louis Putterman & Avner Ben-Ner, pp. 337–63. Cambridge University Press. [HG]

Fehr, E. & Gächter, S. (2000) Cooperation and punishment in public goods experiments. *American Economic Review* 90:980–94. [aMEMcC]

Fehr, E., & Gächter, S. (2002) Altruistic punishment in humans. *Nature* 415:137–40. [RJC]

Fehr, E. & List, J. A. (2004) The hidden costs and returns of incentives – Trust and trustworthiness among CEOs. *Journal of the European Economic Association* 2:743–71. [aMEMcC]

Fehr, R., Gelfand, M. J. & Nag, M. (2010) The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin* 136:894–914. [arMEMcC]

Felson, R. B. (1982) Impression management and the escalation of aggression and violence. *Social Psychology Quarterly* 45:245–54. [aMEMcC]

Fessler, D. M. T. (2004) Shame in two cultures: Implications for evolutionary approaches. *Journal of Cognition and Culture* 4:207–62. [CH]

Fessler, D. M. T. (2006) Violent response to transgression as an example of the intersection of evolved psychology and culture. In: *Missing the revolution: Darwinism for social scientists*, ed. J. Barkow, pp. 101–17. Oxford University Press. [CH]

Fessler, D. M. T. (2010) Madmen: An evolutionary perspective on anger and men's violent responses to transgression. In: *International handbook of anger: Constituent and concomitant biological, psychological, and social processes*, ed. M. Potegal, G. Stemmler & C. Spielberger, pp. 361–81. Springer. [CH, aMEMcC]

Fessler, D. M. T. & Gervais, M. (2010) From whence the captains of our lives: Ultimate and phylogenetic perspectives on emotions in humans and other primates. In: *Mind the gap: The origins of human universals*, ed. P. Kappeler & J. B. Silk, pp. 261–80. Springer. [CH]

Finkel, E. J. & Campbell, W. K. (2001) Self-control and accommodation in close relationships: An interdependence analysis. *Journal of Personality and Social Psychology* 81:263–77. [JCK]

Finkel, E. J., Rusbult, C. E., Kumashiro, M. & Hannon, P. A. (2002) Dealing with betrayal in close relationships: Does commitment promote forgiveness? *Journal of Personality and Social Psychology* 82(6):956–74. [JCK, aMEMcC]

Fiske, A. & Tetlock, P. E. (1997) Taboo trade-offs: Reactions to transactions that transgress spheres of justice. *Political Psychology* 18:255–97. [DL]

Fitness, J. & Peterson, J. (2008) Punishment and forgiveness in close relationships: An evolutionary, social-psychological perspective. In: *Social relationships: Cognitive, affective, and motivational perspectives*, ed. J. P. Forgas & J. Fitness, pp. 255–69. Psychology Press. [aMEMcC]

Foster, J., Shrira, I. & Campbell, W. K. (2006) Theoretical models of narcissism, sexuality, and relationship commitment. *Journal of Social and Personal Relationships* 23(3):367–86. [SK]

Frank, R. H. (1988) *Passions within reason: The strategic role of the emotions*. Norton. [HG, arMEMcC]

Frankenhuis, W. E. & Del Giudice, M. (2012) When do adaptive developmental mechanisms yield maladaptive outcomes? *Developmental Psychology* 48:628–42. [rMEMcC]

Fraser, O. N. & Bugnyar, T. (2011) Ravens reconcile after aggressive conflicts with valuable partners. *PLoS One* 6:1–5. [aMEMcC]

Fraser, O. N., Stahl, D. & Aureli, F. (2010) The function and determinants of reconciliation in *Pan troglodytes*. *International Journal of Primatology* 31:39–57. [aMEMcC]

Frean, M. R. (1994) The prisoner's dilemma without synchrony. *Proceedings of the Royal Society of London, Series B – Biological Sciences* 257:75–79. [aMEMcC]

Freedman, S. R. & Enright, R. D. (1996) Forgiveness as an intervention goal with incest survivors. *Journal of Consulting and Clinical Psychology* 64:983–92. [aMEMcC]

Freyd, J. J. (1996) *Betrayal trauma: The logic of forgetting childhood abuse*. Harvard University Press. [SJ-F]

Freyd, J. J., DePrince, A. P. & Gleaves, D. (2007) The state of betrayal trauma theory: Reply to McNally (2007) – Conceptual issues and future directions. *Memory* 15:295–311. [SJ-F]

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P. & Hewitt, J. K. (2008) Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General* 137:201–25. [JCK]

Frijda, N. H. (1986) *The emotions*. Cambridge University Press. [FA]

Frijda, N. H. (1994) The lex talionis: On vengeance. In: *Emotions: Essays on emotion theory*, ed. S. H. M. van Goozen, N. E. Van de Poll & J. A. Sargeant, pp. 263–89. Erlbaum. [aMEMcC]

Fudenberg, D., Levine, D. K. & Maskin, E. (1994) The folk theorem with imperfect public information. *Econometrica* 62:997–1039. [HG]

Gaertner, S. L. & Dovidio, J. F. (1986) The aversive form of racism. In: *Prejudice, discrimination, and racism*, ed. J. F. Dovidio & S. L. Gaertner, pp. 61–89. Academic Press. [ELU]

Gallup, A., White, D. & Gallup, G. (2007) Handgrip strength predicts sexual behavior, body morphology, and aggression in male college students. *Evolution and Human Behavior* 28:423–29. [aMEMcC]

Gardner, A. & West, S. A. (2004) Cooperation and punishment, especially in humans. *American Naturalist* 164:753–64. [aMEMcC]

Gaylord, M. S. & Galligher, J. F. (1994) Death penalty politics and symbolic law in Hong Kong. *International Journal of the Sociology of Law* 22:19–37. [aMEMcC]

Geary, D. C. (2010) *Male, female: The evolution of human sex differences*. American Psychological Association. [rMEMcC]

Geary, D. C. & Bjorklund, D. F. (2000) Evolutionary developmental psychology. *Child Development* 71:57–65. [rMEMcC]

Gervais, M. M. & Fessler, D. M. T. (under review) You're nothing to me: Reconceptualizing contempt as a cultural model of an ancient attitude. [CH]

Giesbrecht, G. F., Miller, M. R. & Muller, U. (2010) The anger-distress model of temper tantrums: Associations with emotional reactivity and emotional competence. *Infant and Child Development* 19:478–97. [MP]

Gigerenzer, G., Todd, P. M. & The ABC Research Group. (1999) *Simple heuristics that make us smart*. Oxford University Press. [FA, MBP]

Gintis, H. (2000) Strong reciprocity and human sociality. *Journal of Theoretical Biology* 206:169–79. [HG]

Gintis, H. (2003) The hitchhiker's guide to altruism: Genes, culture, and the internalization of norms. *Journal of Theoretical Biology* 220(4):407–18. [HG]

Gintis, H. (2009) *The bounds of reason: Game theory and the unification of the behavioral sciences.* Princeton University Press. [HG]

Gintis, H., Bowles, S., Boyd, R. & Fehr, E. (2005) *Moral sentiments and material interests: On the foundations of cooperation in economic life.* MIT Press. [HG]

Gold, G. J. & Weiner, B. (2000) Remorse, confession, group identity, and expectancies about repeating a transgression. *Basic and Applied Social Psychology* 22:291–300. [aMEMcC]

Goldberg, J. H., Lerner, J. S. & Tetlock, P. E. (1999) Rage and reason: The psychology of the intuitive prosecutor. *European Journal of Social Psychology* 29:781–95. [KO]

Gopnik, A. (1996) Theories and modules; Creation myths, developmental realities, and Neurath's boat. In: *Theories of theories of mind*, ed. P. Carruthers & P. K. Smith, pp. 169–83. Cambridge University Press. [TJW]

Gottlieb, G. (2007) Probabilistic epigenesis. *Developmental Science* 10:1–11. [TJW]

Gould, R. V. (2000) Revenge as sanction and solidarity display: An analysis of vendettas in nineteenth-century Corsica. *American Sociological Review* 65:682–704. [aMEMcC]

Govier, T. (2002) *Forgiveness and revenge*. Routledge. [aMEMcC]

Graham, J., Haidt, J. & Nosek, B. A. (2009) Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology* 96:1029–46. [KO]

Graham, K. & Wells, S. (2003) 'Somebody's Gonna get their head kicked in tonight!' Aggression among young males in bars-a question of values? *The British Journal of Criminology* 43:546–66. [MP]

Gray, K., Young, L. & Waytz, A. (in press) Mind perception is the essence of morality. *Psychological Inquiry*. [rMEMcC]

Green, J. A., Whitney, P. G. & Potegal, M. (2011) Screaming, yelling, whining, and crying: Categorical and intensity differences in vocal expressions of anger and sadness in children's tantrums. *Emotion* 11:1124–33. [MP]

Greenwald, A. G. & Banaji, M. R. (1995) Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review* 102(1):4–27. [RF]

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A. & Mellott, D. S. (2002) A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review* 109(1):3–25. [RF]

Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. (1998) Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology* 74(6):1464–480. [RF]

Griskevicius, V., Tybur, J., Delton, A. & Robertson, T. (2011) The influence of mortality and socioeconomic status on risk and delayed rewards: A life history theory approach. *Journal of Personality and Social Psychology* 100:1015–26. [rMEMcC]

Guala, F. (2012) Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences* 35(1):1–59. [rMEMcC, DR]

Gullo, M. J. & Dawe, S. (2008) Impulsivity and adolescent substance use: Rashly dismissed as "all-bad"? *Neuroscience and Biobehavioral Reviews* 32:1507–28. [rMEMcC]

Gürerk, O., Irlenbusch, B. & Rockenbach, B. (2006) The competitive advantage of sanctioning institutions. *Science* 312:108–11. [aMEMcC]

Güth, W., Schmittberger, R. & Schwarze, B. (1982) An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3:367–88. [aMEMcC]

Hagen, E. H. & Hammerstein, P. (2006) Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology* 69:339–48. [arMEMcC]

Haidt, J. (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108:814–34. [KO, ELU]

Haidt, J., Koller, S. & Dias, M. (1993) Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology* 65:613–28. [ELU]

Hamilton, W. D. (1964) The genetical evolution of social behaviour, I & II. *Journal of Theoretical Biology* 7(1):1–52. [aMEMcC]

Harbaugh, W. T. (1998) The prestige motive for making charitable transfers. *American Economic Review, Papers and Proceedings* 88:277–82. [CH]

Harcourt, A. H. & de Waal, F. B. M., eds. (1992) *Coalitions and alliances in humans and other animals*. Oxford University Press. [rMEMcC]

Harris, J. R. (2006) *No two alike: Human nature and human individuality*. Norton. [aMEMcC]

Haselton, M. G. & Buss, D. M. (2000) Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology* 78(1):81–91. [PB]

Hauert, C. & Schuster, H. G. (1998) Extending the iterated prisoner's dilemma without synchrony. *Journal of Theoretical Biology* 192:155–66. [aMEMcC]

Hawkes, K. (1993) Why hunter-fatherers work: An ancient version of the problem of public goods. *Current Anthropology* 34(4):341–61. [HG]

Hayashi, A., Abe, N., Ueno, A., Shigemune, Y., Mori, E., Tashiro, M. & Fujii, T. (2010) Neural correlates of forgiveness for moral transgressions involving deception. *Brain Research* 1332:90–99. doi: 10.1016/j.brainres.2010.03.045 [RF]

Hayashi, N., Ostrom, E., Walker, J. & Yamagishi, T. (1999) Reciprocity, trust, and the sense of control: A cross-societal study. *Rationality and Society* 11:27–46. [aMEMcC]

Hebl, J. & Enright, R. D. (1993) Forgiveness as a psychotherapeutic goal with elderly females. *Psychotherapy: Theory, Research, Practice, Training* 30:658–67. [aMEMcC]

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E. & Gintis, H. (2004) *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford University Press. [HG]

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N., Hill, K., Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J. Q. & Tracer, D. (2005) "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences* 28(6):795–815. [HG]

Henrich, J. (2009) The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and Human Behavior* 30:244–60. [aMEMcC]

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F. W., Tracer, D. & Ziker, J. (2006) Costly punishment across human societies. *Science* 312:1767–70. [HG, aMEMcC]

Herrmann, B., Thöni, C. & Gächter, S. (2008) Antisocial punishment across societies. *Science* 319:1362–67. [aMEMcC]

Hess, N. H. & Hagen, E. H. (2006) Sex differences in indirect aggression: Psychological evidence from young adults. *Evolution and Human Behavior* 27:231–45. [aMEMcC]

Hewstone, M., Cairns, E., Voci, A., Hamberger, J. & Niens, U. (2006) Intergroup contact, forgiveness, and the experience of "The troubles" in Northern Ireland. *Journal of Social Issues* 6:99–120. [RJC]

Hill, E. M. & Chow, K. (2006) Life-history theory and risky drinking. *Addiction* 97:401–13. [rMEMcC]

Hill, E. M., Jenkins, J. & Farmer, L. (2008) Family unpredictability, future discounting, and risk taking. *Journal of Socio-Economics* 37:1381–96. [rMEMcC]

Hinde, R. A. (1979) *Towards understanding relationships*. Academic Press. [FA]

Hirshleifer, J. (1987) On the emotions as guarantors of threats and promises. In: *The latest on the best: Essays on evolution and optimality*, ed. J. Dupré, pp. 307–26. MIT Press. [aMEMcC]

Ho, R., ForsterLee, L., ForsterLee, R. & Crofts, N. (2002) Justice versus vengeance: Motives underlying punitive judgements. *Personality and Individual Differences* 33:365–77. [SCR]

Hokanson, J. E. & Burgess, M. (1962) The effects of status, type of frustration, and aggression on vascular processes. *Journal of Abnormal and Social Psychology* 65:232–37. [VJK]

Hokanson, J. E., Burgess, M. & Cohen, M. F. (1963) Effects of displaced aggression on systolic blood pressure. *Journal of Abnormal and Social Psychology* 67:214–18. [VJK]

Hokanson, J. E. & Shetler, S. (1961) The effect of overt aggression on physiological arousal level. *Journal of Abnormal and Social Psychology* 63:446–48. [VJK]

Holtzman, N. S. & Strube, M. J. (2010) Narcissism and attractiveness. *Journal of Research in Personality* 44:133–36. [rMEMcC]

Hoover, J. P. & Robinson, S. K. (2007) Retaliatory mafia behavior by a parasitic cowbird favors host acceptance of parasitic eggs. *Proceedings of the National Academy of Sciences USA* 104:4479–83. [arMEMcC]

Hope, D. (1987) The healing paradox of forgiveness. *Psychotherapy: Theory, Research, Practice, Training* 24:240–44. [aMEMcC]

Hopfensitz, A. & Reuben, E. (2009) The importance of emotions for the effectiveness of social punishment. *The Economic Journal* 119:1534–59. [aMEMcC]

Horney, K. (1948) The value of vindictiveness. *American Journal of Psychoanalysis* 8:3–12. [aMEMcC]

Hoyt, W. T., Fincham, F., McCullough, M. E., Maio, G. & Davila, J. (2005) Responses to interpersonal transgressions in families: Forgivingness, forgivability, and relationship-specific effects. *Journal of Personality and Social Psychology* 89:375–94. [rMEMcC]

Hruschka, D. J. & Henrich, J. (2006) Friendship, cliquishness, and the emergence of cooperation. *Journal of Theoretical Biology* 239:1–15. [aMEMcC]

Hsee, C. K., Loewenstein, G. F., Blount, S. & Bazerman, M. H. (1999) Preference reversals between joint and separate evaluation of options: A review and theoretical analysis. *Psychological Bulletin* 125:576–90. [ELU]

Human Development Study Group (1991) Five points on the construct of forgiveness within psychotherapy. *Psychotherapy: Theory, Research, Practice, Training* 28:493–96. [aMEMcC]

Inbar, Y., Pizarro, D., Iyer, R. & Haidt, J. (2012) Disgust sensitivity, political conservatism, and voting. *Social Psychological and Personality Science* 5:537–44. [KO]

Ingle, D. (2004) Recreational fighting. In: *Encyclopedia of recreation and leisure in America, vol. 2*, ed. G. S. Cross, pp. 198–200. Scribner. [MP]

Ishikawa, S. S., Raine, A., Lencz, T., Bihrle, S. & LaCasse, L. (2001) Increased height and bulk in antisocial personality disorder and its subtypes. *Psychiatry Research* 105:211–19. [aMEMcC]

Jacoby, S. (1983) *Wild justice: The evolution of revenge*. Harper and Row. [aMEMcC]

Jensen, K. (2010) Punishment and spite, the dark side of cooperation. *Philosophical Transactions of the Royal Society B* 365:2635–50. [aMEMcC]

Jensen, K., Call, J. & Tomasello, M. (2007) Chimpanzees are vengeful but not spiteful. *Proceedings of the National Academy of Sciences USA* 104:13046–50. [arMEMcC]

Johnson, M. (1993) Moral imagination: Implications of cognitive science for ethics. University of Chicago Press. [DJS]

Johnson-Laird, P. N. & Oatley, K. (1992) The language of emotions: An analysis of a semantic field. *Cognition and Emotion* 3:81–123. [FA]

Johnson-Laird, P. N. & Oatley, K. (1992) Basic emotions, rationality, and folk theory. *Cognition and Emotion* 6:201–23. [FA]

Joskowicz-Jabloner, L. & Leiser, D. (forthcoming) Varieties of trust-betrayal: Emotion and relief patterns in different domains. *Journal of Applied Social Psychology*. Available at SSRN: http://ssrn.com/abstract=1955637 [DL]

Judge, P. (1982) Redirection of aggression based on kinship in a captive group of pigtail macaques. *International Journal of Primatology* 3:301. [FA]

Kääriäinen, J. T. (2007) Trust in the police in 16 European societies: A multilevel analysis. *European Journal of Criminology* 4:409–35. [aMEMcC]

Kachadourian, L. K., Fincham, F. & Davila, J. (2004) The tendency to forgive in dating and married couples: The role of attachment and relationship satisfaction. *Personal Relationships* 11:373–93. [aMEMcC]

Kahneman, D., Schkade, D. & Sunstein, C. (1998) Shared outrage and erratic awards: The psychology of punitive damages. *Journal of Risk and Uncertainty* 16:49–86. [KO]

Kaminer, D, Stein, D. J., Mbanga, I. & Zungu-Dirwayi, N. (2001) The Truth and Reconciliation Commission in South Africa: Relation to psychiatric status and forgiveness among survivors of human rights abuses. *The British Journal of Psychiatry; The Journal of Mental Science* 178:373–77. [DJS]

Kaplan, H., Hill, K., Hawkes, K. & Hurtado, A. M. (1984) Food sharing among Ache hunter-gatherers of Eastern Paraguay. *Current Anthropology* 25(1):113–15. [HG]

Kappeler, P. M. (1993) Reconciliation and post-conflict behaviour in ringtailed lemurs, *Lemur catta* and redfronted lemurs, *Eulemur fulvus rufus*. *Animal Behaviour* 45(5):901–15. [AD]

Karremans, J. C. & Aarts, H. (2007) The role of automaticity in determining the inclination to forgive close others. *Journal of Experimental Social Psychology* 43:902–17. [JCK, aMEMcC]

Karremans, J. C., Regalia, C., Paleari, F. G., Finham, F. D., Cui, M., Takada, N., Ohbuchi, K., Terzino, K., Cross, S. E. & Uskul, A. K. (2011) Maintaining harmony across the globe: The cross-cultural association between closeness and interpersonal forgiveness. *Social Psychology and Personality Science* 2:443–51. [JCK]

Katsukake, N. & Castles, D. L. (2001) Reconciliation and variation in post-conflict stress in Japanese macaques (*Macaca fuscata fuscata*): Testing the integrated hypothesis. *Animal Cognition* 4:259–68. [aMEMcC]

Kearns, J. N. & Fincham, F. D. (2005) Victim and perpetrator accounts of interpersonal transgressions: Self-serving or relationship-serving biases? *Personality and Social Psychology Bulletin* 31(3):321–33. [SK]

Keltner, D. & Buswell, B. N. (1997) Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin* 122:250–70. [aMEMcC]

Kim, S. H., Smith, R. H. & Brigham, N. L. (1998) Effects of power imbalance and the presence of third parties on reactions to harm: Upward and downward revenge. *Personality and Social Psychology Bulletin* 24:353–61. [aMEMcC]

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P. & Montague, P. R. (2008) The rupture and repair of cooperation in borderline personality disorder. *Science* 321:806–10. [aMEMcC]

Klapwijk, A. & Van Lange, P. A. M. (2009) Promoting cooperation and trust in noisy situations: The power of generosity. *Journal of Personality and Social Psychology* 96:83–103. [aMEMcC]

Knauft, B. M. (1987) Reconsidering violence in simple human societies: Homicide among the Gebusi of New Guinea. *Current Anthropology* 28:457–500. [MP]

Koehler, J. J. & Gershoff, A. D. (2003) Betrayal aversion: When agents of protection become agents of harm. *Organizational Behavior and Human Decision Processes* 90:244–61. [DL]

Koenig, B. L. & Ketelaar, T. (2006) Why big guys don't smile: Preliminary evidence for reactive heritability in collegiate football players. Paper presented at the Human Behavior and Evolution Society Meeting, Philadelphia, PA, June 2006. [aMEMcC]

Kollock, P. (1998) Social dilemmas: Anatomy of cooperation. *Annual Review of Sociology* 24:183–214. [RJC, aMEMcC]

Konečni, V. J. (1975a) Annoyance, type and duration of postannoyance activity, and aggression: The "cathartic effect." *Journal of Experimental Psychology: General* 104:76–102. [VJK]

Konečni, V. J. (1975b) The mediation of aggressive behavior: Arousal level versus anger and cognitive labeling. *Journal of Personality and Social Psychology* 32:706–12. [VJK]

Konečni, V. J. (1984) Methodological issues in human aggression research. In: *Aggression in children and youth*, ed. R. M. Kaplan, V. J. Konečni & R. W. Novaco, pp. 1–43. Martinus Nijhoff. [VJK]

Konečni, V. J., Crozier, J. B. & Doob, A. N. (1976) Anger and expression of aggression: Effects on aesthetic preference. *Scientific Aesthetics* 1:47–55. [VJK]

Konečni, V. J. & Doob, A. N. (1972) Catharsis through displacement of aggression. *Journal of Personality and Social Psychology* 23:379–87. [VJK]

Konečni, V. J. & Ebbesen, E. B. (1976) Disinhibition versus the cathartic effect: Artifact and substance. *Journal of Personality and Social Psychology* 34:352–65. [VJK]

Konrath, S., Bushman, B. & Grove, T. (2009) Seeing my world in a million little pieces: Narcissism, self-construal, and cognitive-perceptual style. *Journal of Personality* 77(4):1197–228. doi: 10.1111/j.1467-6494.2009.00579.x [SK]

Koski, S. E., Koops, K. & Sterck, E. H. M. (2007) Reconciliation, relationship quality, and postconflict anxiety: Testing the integrated hypothesis in captive chimpanzees. *American Journal of Primatology* 69:158–72. [aMEMcC]

Koyama, N. F. (2001) The long-term effects of reconciliation in Japanese macaques, *Macaca fuscata*. *Ethology* 107:975–87. [aMEMcC]
Kreps, D. M. & Wilson, R. (1982) Sequential equilibria. *Econometrica* 50(4):863–94. [HG]
Kubrin, C. E. & Weitzer, R. (2003) Retaliatory homicide: Concentrated disadvantage and neighborhood culture. *Social Problems* 50:157–80. [aMEMcC]
Kuhl, J. & Koole, S. L. (2004) Workings of the will: A functional approach. In: *Handbook of experimental existential psychology*, ed. J. Greenberg, S. L. Koole & T. Pyszczynski, pp. 411–30. Guildford Press. [FP]
Kurzban, R. (2010a) Does the brain consume additional glucose during self-control tasks. *Evolutionary Psychology* 8:244–59. [rMEMcC]
Kurzban, R. (2010b) *Why everyone (else) is a hypocrite: Evolutionary psychology and the modular mind*. Princeton University Press. [rMEMcC]
Kurzban, R. & DeScioli, P. (submitted) Adaptationist punishment in humans. [aMEMcC]
Kurzban, R., DeScioli, P. & Fein, D. (2012) Hamilton vs. Kant: Pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior* 32(4):323–33. [rMEMcC]
Kurzban, R., DeScioli, P. & O'Brien, E. (2007) Audience effects on moralistic punishment. *Evolution and Human Behavior* 28:75–84. [rMEMcC]
Kurzban, R., Tooby, J. & Cosmides, L. (2001) Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences USA* 98:15387–92. [rMEMcC]
Lachs, M. & Pillemer, K. (2004) Elder abuse. *Lancet* 364:1263–72. [ANS]
Lapsley, J. N. (1998) Vengeance: The half-hidden pillager of our lives. *Pastoral Psychology* 46:255–65. [SCR]
Lassek, W. D. & Gaulin, S. J. C. (2009) Costs and benefits of fat-free muscle mass in men: Relationship to mating success, dietary requirements, and native immunity. *Evolution and Human Behavior* 30:322–28. [aMEMcC]
Lazarus, R. (1991) *Emotion and adaptation*. Oxford University Press. [CH]
LeDoux, J. E. (1996) The emotional brain. Simon & Schuster. [FA]
Ledyard, J. O. (1995) Public goods: A survey of experimental research. In: *Handbook of experimental economics*, ed. J. Kagel & A. E. Roth, pp. 111–94. Princeton University Press. [aMEMcC]
Lennon, R. & Eisenberg, N. (1989) Gender and age differences in emapthy and sympathy. In: *Empathy and its development*, ed. N. Eisenberg & J. Strayer, pp. 195–217. Cambridge University Press. [SK]
Lerner, J. S., Goldberg, J. H. & Tetlock, P. E. (1998) Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin* 24:563–74. [KO]
Lerner, M. J. (1980) *The belief in a just world: A fundamental delusion*. Plenum Press. [FP]
Lerner, M. J. (1987) Integrating societal and psychological rules of entitlement: The basic task of each social actor and fundamental problem for the social sciences. *Social Justice Research* 1:107–25. [FP]
Lerner, M. J. (2003) The justice motive: Where social psychologists found it, how they lost it, and why they may not find it again. *Personality and Social Psychology Review* 7:388–99. [SCR]
Lerner, M. J. & Miller, D. T. (1978) Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin* 85:1030–51. [FP]
Lerner, M. J., Miller, D. T. & Holmes, J. G. (1976) Deserving and the emergence of forms of justice. In: *Advances in experimental social psychology, vol. 9*, ed. L. Berkowitz & E. Walster, pp. 133–62. Academic Press. [FP]
Leslie, A. M. (1987) Pretense and representation: The origins of "theory of mind." *Psychological Review* 94:412–26. [TJW]
LeVine, R. A. & Campbell, D. T. (1972) *Ethnocentrism: Theories of conflict, ethnic attitudes and group behaviour*. Wiley. [RJC]
Liberman, N., Trope, Y. & Stephan, E. (2007) Psychological distance. In: *Social psychology: Handbook of basic principles*, ed. A. W. Kruglanski & E. T. Higgins, Guilford Press. [FP]
Lickel, B., Miller, N., Stenstrom, D. M., Denson, T. F. & Schmader, T. (2006) Vicarious retribution: The role of collective blame in intergroup aggression. *Personality and Social Psychology Review* 10:372–90. [aMEMcC]
Lieberman, D. & Linke, L. (2007) The effect of social category on third party punishment. *Evolutionary Psychology* 5:289–305. [aMEMcC, MBP]
Lieberman, D., Tooby, J. & Cosmides, L. (2007) The architecture of human kin detection. *Nature* 445:727–31. [aMEMcC]
Loewenstein, G. (1996) Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes* 65:272–92. [KO]
Lott, D. F. (1991) *Intraspecific variation in the social system of wild vertebrates*. Cambridge University Press. [FA]
Lundahl, B. W., Taylor, M. J., Stevenson, R. & Roberts, K. D. (2008) Process-based forgiveness interventions: A meta-analytic review. *Research on Social Work Practice* 18:465–78. [aMEMcC]
Macaskill, A., Maltby, J. & Day, L. (2002) Forgiveness of self and others and emotional empathy. *The Journal of Social Psychology* 142(5):663–65. [SK]
Maguire, E. A. (2001) Neuroimaging studies of autobiographical event memory. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 356(1413):1441–51. [RF]
Marks, M. J. (1988) Remorse, revenge, and forgiveness. *Psychotherapy Patient* 5:317–30. [aMEMcC]
Marlatt, G. A., Kosturn, C. F. & Lang, A. R. (1975) Provocation to anger and opportunity for retaliation as determinants of alcohol consumption in social drinkers. *Journal of Abnormal Psychology* 84:652–59. [VJK]
Marlowe, F. W., Berbesque, J. C., Barrett, C., Bolyanatz, A., Gurven, M. & Tracer, D. (2011) The "spiteful" origins of human cooperation. *Proceedings of the Royal Society of London B* 278:2159–64. [aMEMcC]
Matsumura, S. & Hayden, T. J. (2006) When should signals of submission be given? – A game theory model. *Journal of Theoretical Biology* 240:425–33. [aMEMcC]
May, M. E. & Kennedy, C. H. (2009) Aggression as positive reinforcement in mice under various ratio-and time-based reinforcement schedules. *Journal of Experimental Analysis of Behavior* 91:185–96. [MP]
Maynard Smith, J. & Harper, D. (2003) *Animal signals*. Oxford. [aMEMcC]
McCorkle, R. C. (1993) Research note: Punish and rehabilitate? Public attitudes toward six common crimes. *Crime and Delinquency* 39(2):240–52. [MBP]
McCullough, M. E. (2001) Forgiveness: Who does it and how do they do it? *Current Directions in Psychological Science* 10, 194–97. [rMEMcC]
McCullough, M. E. (2008) *Beyond revenge: The evolution of the forgiveness instinct*. Jossey-Bass. [arMEMcC]
McCullough, M. E., Bellah, C. G., Kilpatrick, S. D. & Johnson, J. L. (2001) Vengefulness: Relationships with forgiveness, rumination, well-being, and the Big Five. *Personality and Social Psychology Bulletin* 27:601–10. [arMEMcC]
McCullough, M. E., Emmons, R. A., Kilpatrick, S. D. & Mooney, C. N. (2003) Narcissists as "victims": The role of narcissism in the perception of transgressions. *Personality and Social Psychology Bulletin* 29(7):885–93. [SK]
McCullough, M. E., Fincham, F. D. & Tsang, J. (2003) Forgiveness, forbearance, and time: The temporal unfolding of transgression-related interpersonal motivations. *Journal of Personality and Social Psychology,* 84:540–57. [aMEMcC]
McCullough, M. E. & Hoyt, W. T. (2002) Transgression-related motivational dispositions: Personality substrates of forgiveness and their links to the Big Five. *Personality and Social Psychology Bulletin* 28:1556–73. [arMEMcC]
McCullough, M. E., Luna, L. R., Berry, J. W., Tabak, B. A. & Bono, G. (2010) On the form and function of forgiving: Modeling the time-forgiveness relationship and testing the valuable relationships hypothesis. *Emotion* 3:358–76. [aMEMcC]
McCullough, M. E., Pedersen, E. J., Schroder, J., Tabak, B. & Carver, C. S. (2012) Childhood exposure to neglect, conflict, and crime predict unprovoked and retaliatory defection in an iterated prisoner's dilemma. Unpublished manuscript. University of Miami. [rMEMcC]
McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L., Jr., Brown, S. W. & Hight, T. L. (1998) Interpersonal forgiving in close relationships. II: Theoretical elaboration and measurement. *Journal of Personality and Social Psychology* 75:1586–603. [aMEMcC]
McCullough, M. E. & Root, L. M. (2005) Forgiveness as change. In: *Handbook of forgiveness*, ed. E. L. Worthington, Jr., pp. 91–107. Routledge. [aMEMcC]
McCullough, M. E. & Worthington, E. L., Jr. (1995) Promoting forgiveness: A comparison of two brief psychoeducational group interventions with a waiting-list control. *Counseling and Values* 40:55–68. [aMEMcC]
McCullough, M. E. & Worthington, E. L., Jr. (1999) Religion and the forgiving personality. *Journal of Personality* 67:1141–64. [rMEMcC]
McCullough, M. E., Worthington, E. L., Jr. & Rachal, K. C. (1997) Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology* 73:321–36. [aMEMcC]
McDermott, R., Tingley, D., Cowden, J., Frazzetto, G. & Johnson, D. D. P. (2009) Monoamine oxidase A gene (MAOA) predicts behavioral aggression following provocation. *Proceedings of the National Academy of Sciences USA* 106:2118–23. [aMEMcC]
Miller, A. J., Worthington, E. L., Jr. & McDaniel, M. A. (2008) Forgiveness and gender: A meta-analytic review and research agenda. *Journal of Social and Clinical Psychology* 27:843–76. [aMEMcC]
Mobbs, D., Yu, R., Meyer, M., Passamonti, L., Seymour, B., Calder, A. J., Schweizer, S., Frith, C. D. & Dalgleish, T. (2009) A key role for similarity in vicarious reward. *Science* 324(5929):900. [RY]
Mocan, N. H. (2008) *Vengeance*. Working Paper, National Bureau of Economic Research, Cambridge, MA, June 2008. Paper No, 14131. National Bureau of Economic Research. Retrieved December 20, 2011 from: http://www.nber.org/papers/w14131 [aMEMcC]
Molden, D. C., Hui, C. M., Scholer, A. A., Meier, B. P., Noreen, E. E., D'Agostino, P. R. & Martin, V. (in press) The motivational versus metabolic effects of carbohydrates on self-control. *Psychological Science*. [rMEMcC]
Molm, L. D. (1997) *Coercive power in social exchange*. Cambridge University Press. [aMEMcC]

Montague, P. R. & Berns, G. S. (2002) Neural economics and the biological substrates of valuation. *Neuron* 36(2):265–84. [AD]

Moss, D. B. (1986) Revenge and forgiveness. *American Imago* (43):191–210. [aMEMcC]

Mullet, E., Neto, F. & Riviere, S. (2005) Personality and its effects on resentment, revenge, forgiveness, and self-forgiveness. In: *Handbook of forgiveness*, ed. E. L. Worthington, Jr., pp. 159–81. Routledge/Taylor & Francis. [SK, aMEMcC]

Murphy, J. G. (2003) *Getting even: Forgiveness and its limits*. Oxford. [aMEMcC]

Murray, J., Thomson, M. E., Cooke, D. J. & Charles, K. E. (2011) Investigating the relationship between justice-vengeance motivations in punitive sentencing recommendations. *Legal and Criminological Psychology* [Published Online, ahead of print]. doi: 10.1111/j.2044-8333.2011.02021.x. [SCR]

Nadler, A. & Liviatan, I. (2006) Intergroup reconciliation: Effects of adversary's expressions of empathy, responsibility, and recipients' trust. *Personality and Social Psychology Bulletin* 32:459–70. [aMEMcC]

Navon, D. (1984) Resources-a theoretical soup stone? *Psychological Review* 91:216–34. [rMEMcC]

Nelissen, M. H. J. (2008) The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior* 29:242–48. [aMEMcC]

Nell, V. (2006) Cruelty's rewards: The gratifications of perpetrators and spectators. *Behavioral and Brain Sciences* 29:211–57. [MP]

Nelson, J. D. (2005) Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review* 112:979–99. [ELU]

Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W. & Sejnowski, T. J. (2010) Experience matters: Information acquisition optimizes probability gain. *Psychological Science* 21:960–69. [ELU]

Nesse, R. M. (2005) Natural selection and the regulation of defenses: A signal detection analysis of the smoke detector principle. *Evolution and Human Behavior* 26:88–105. [PB]

Nesse, R. M. & Williams, G. C. (1994) *Why we get sick: The new science of Darwinian medicine*. Vintage Books. [aMEMcC]

Nettle, D. (2006) The evolution of personality variation in humans and other animals. *American Psychologist* 20:622–31. [rMEMcC]

Nettle, D. (2009) Ecological influences on human behavioural diversity: A review of recent findings. *Trends in Ecology and Evolution* 24:618–24. [rMEMcC]

Nikiforakis, N. & Normann, H. (2008) A comparative statics analysis of punishment in public goods experiments. *Experimental Economics* 11:358–69. [aMEMcC]

Nisbett, R. E. & Cohen, D. (1996) *Culture of honor: The psychology of violence in the South*. Westview. [aMEMcC]

Nisbett, R. E. & Wilson, T. D. (1977) Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84(3):231–59. [RF, KO]

Nowak, M. A. & Sigmund, K. (1993) A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma Game. *Nature* 364:56–58. [aMEMcC]

O'Gorman, R., Sloan Wilson, D. & Miller, R. R. (2005) Altruistic punishing and helping differ in sensitivity to relatedness, friendship and future interactions. *Evolution and Human Behavior* 26:375–87. [SCR]

Olson, J. M., Vernon, P. A., Harris, J. A. & Jang, K. L. (2001) The heritability of attitudes: A study of twins. *Journal of Personality and Social Psychology* 80:845–60. [aMEMcC]

Oosterbeek, H., Sloof, R. & Van de Kuilen, G. (2004) Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics* 7:171–88. [aMEMcC]

Orth, U., Berking, M., Walker, N., Meier, L. L. & Znoj, H. (2008) Forgiveness and psychological adjustment following interpersonal transgressions: A longitudinal analysis. *Journal of Research in Personality* 42:365–85. [aMEMcC]

Orth, U., Montada, L. & Maercker, A. (2006) Feelings of revenge, retaliation motive, and posttraumatic stress reactions in crime victims. *Journal of Interpersonal Violence* 21:229–43. [aMEMcC]

Ostrom, E. (1990) *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press. [RJC, aMEMcC]

Ostrom, E., Walker, J. & Gardner, R. (1992) Covenants with and without a sword: Self governance is possible. *American Political Science Review* 86:404–16. [aMEMcC]

Owren, M. J. & Rendall, D. (1997) An affect-conditioning model of nonhuman primate vocal signaling. *Perspectives in Ethology* 12:299–346. [FA]

Page, T., Putterman, L. & Unel, B. (2005) Voluntary association in public goods experiments: Reciprocity, mimicry, and efficiency. *The Economic Journal* 115:1032–53. [aMEMcC]

Palagi, E., Antonaccia, D. & Norscia, I. (2008) Peacemaking on treetops: First evidence of reconciliation from a wild prosimian (*Propithecus verreauxi*) *Animal Behaviour* 76:737–47. [aMEMcC]

Panksepp, J. (1989) The psychobiology of emotions: The animal side of human feelings. In: *Emotions and the dual brain*, ed. G. Gainotti & C. Caltagirone, pp. 31–55. Springer-Verlag. [FA]

Paolini, S., Hewstone, M., Cairns, E. & Voci, A. (2004) Effects of direct and indirect cross-group friendships on judgments of Catholics and Protestants in Northern Ireland: The mediating role of anxiety-reduction mechanism. *Personality and Social Psychology Bulletin* 30:770–86. [RJC]

Parkes, C. M. (1993) Psychiatric problems following bereavement by murder or manslaughter. *British Journal of Psychiatry* 162:49–54. [aMEMcC]

Parsons, T. (1967) *Sociological theory and modern society*. Free Press. [HG]

Patterson, G. R. & Cobb, J. A. (1971) A dyadic analysis of aggressive behaviors. In: *Minnesota symposia on child psychology, vol. 5*, ed. J. P. Hill. University of Minnesota Press. [VJK]

Perunovic, M. & Holmes, J. G. (2008) Automatic accommodation: The role of personality. *Personal Relationships* 15:57–70. doi:10.1111/j.1475-6811.2007.00184.x [DB]

Petersen, M. B. (2009) Public opinion and evolved heuristics: The role of category-based inference. *Journal of Cognition and Culture* 9(3):367–89. [MBP]

Petersen, M. B., Sell, A., Tooby, J. & Cosmides, L. (2010) Evolutionary psychology and criminal justice: A recalibrational theory of punishment and reconciliation. In: *Human morality and sociality: Evolutionary and comparative perspectives*, ed. H. Høgh-Olesen, pp. 72–131. Palgrave MacMillan. [arMEMcC, MBP, ANS]

Petersen, M. B., Sell, A., Tooby, J. & Cosmides, L. (2012) To punish or repair? Evolutionary psychology and lay intuitions about modern criminal justice. *Evolution and Human Behavior*. [Published Online, July 16, 2012.] doi:10.1016/j.evolhumbehav.2012.05.003. [MBP]

Pettigrew, T. F. & Tropp, L. R. (2006) A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology* 90:751–83. [RJC]

Phillips, L. J. & Osborne, J. W. (1989) Cancer patients' experiences of forgiveness therapy. *Canadian Journal of Counselling* 23:236–51. [aMEMcC]

Pigliucci, M. & Müller, G. B. (2010) *Evolution: The extended synthesis*. MIT Press. [TJW]

Pillemer, K. & Finkelhor, D. (1988) The Prevalence of elder abuse: A random sample survey. *The Gerontologist* 28(1):51–57. [ANS]

Pingleton, J. P. (1989) The role and function of forgiveness in the psychotherapeutic process. *Journal of Psychology and Theology* 17:27–35. [aMEMcC]

Pinker, S. (1999) How the mind works. *Annals of the New York Academy of Sciences* 882:119–27. [rMEMcC]

Pizarro, D. A. & Tannenbaum, D. (2011) Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In: *The social psychology of morality: Exploring the causes of good and evil*, ed. P. Shaver & M. Mikulincer, APA Books. [ELU]

Pizarro, D. A. & Uhlmann, E. L. (2005) Do normative standards advance our understanding of moral judgment? *Behavioral and Brain Sciences* 28:558–59. [ELU]

Potegal, M. (2000) Post tantrum affiliation with parents: The ontogeny of reconciliation. In: *Natural conflict resolution*, ed. F. Aureli & F. B. M. de Waal, pp. 253–55. University of California Press. [MP]

Potegal, M. (2010) The temporal dynamics of anger: Phenomena, processes and perplexities. In: *International handbook of anger*, ed. M. Potegal, G. Stemmler & C. Spielberger, pp. 385–401. Springer. [MP]

Potegal, M., Carlson, G., Margulies, D., Gutkovitch, Z. & Wall, M. (2009) Rages or temper tantrums? The behavioral organization, temporal characteristics, and clinical significance of angry-agitated outbursts in child psychiatry inpatients. *Child Psychiatry and Human Development* 40:621–36. [MP]

Potegal, M. & Davidson, R. J. (1997) Young children's post tantrum affiliation with their parents. *Aggressive Behavior* 23:329–42. (Special Issue on *Appeasement and Reconciliation*). [MP]

Potegal, M. & Davidson, R. J. (2003) Temper tantrums in young children: 1. Behavioral composition. *Journal of Developmental and Behavioral Pediatrics* 24:140–47. [MP]

Potegal, M., Kosorok, M. R. & Davidson, R. J. (2003) Temper tantrums in young children: II. Tantrum duration and temporal organization. *Journal of Developmental and Behavioral Pediatrics* 24:148–54. [MP]

Preuschoft, S., Wang, X., Aureli, F. & de Waal, F. B. M. (2002) Reconciliation in captive chimpanzees: A reevaluation with controlled methods. *International Journal of Primatology* 23:29–50. [aMEMcC]

Price, M. D. (1997) Can mediation produce justice? A restorative discussion for mediators. *Signal* 1:1–6. [SCR]

Price, M. E., Cosmides, L. & Tooby, J. (2002) Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior* 23:203–31. [aMEMcC]

Pronk, T. M., Karremans, J. C., Overbeek, G., Vermuist, A. A. & Wigboldus, D. H. J. (2010) What it takes to forgive: When and why executive functioning facilitates forgiveness. *Journal of Personality and Social Psychology* 98(1):119–31. [DB, RF, JCK, aMEMcC]

Pryce, C. R. (1996) Socialization, hormones, and the regulation of maternal behaviour in nonhuman simian primates. *Advances in the Study of Behavior* 25:423–73. [FA]

Ramírez, J. M., Bonniot-Cabanac, M.-C. & Cabanac, M. (2005) Can impulsive aggression provide pleasure? *European Psychologist* 10:449–72. [MP]

Rand, D. G., Ohtsuki, H. & Nowak, M. A. (2009) Direct reciprocity with costly punishment: Generous tit-for-tat prevails. *Journal of Theoretical Biology* 256:45–57. [aMEMcC]

Rapoport, A. & Chammah, A. M. (1965) *The prisoner's dilemma*. University of Michigan Press. [aMEMcC]

Reeder, G. D. & Brewer, M. B. (1979) A schematic model of dispositional attribution in interpersonal perception. *Psychological Review* 86:61–79. [ELU]

Rhodewalt, F. & Morf, C. C. (1998) On self-aggrandizement and anger: A temporal analysis of narcissism and affective reactions to success and failure. *Journal of Personality and Social Psychology* 74(3):672–85. [SK]

Richard, F. D., Bond, C. F., Jr. & Stokes-Zoota, J. J. (2003) One hundred years of social psychology quantitatively described. *Review of General Psychology* 7:331–63. [aMEMcC]

Riek, B. M. & Mania, E. W. (2012) The antecedents and consequences of interpersonal forgiveness: A meta-analytic review. *Personal Relationships* 19:304–25. [rMEMcC]

Risen, J. L. & Gilovich, T. (2007) Target and observer differences in the acceptance of questionable apologies. *Journal of Personality and Social Psychology* 92:418–33. [aMEMcC]

Ritzman, T. A. (1987) Forgiveness: Its role in therapy. *Medical Hypnoanalysis Journal* 2:4–13. [aMEMcC]

Roberts, J. V. (1992) Public opinion, crime, and criminal justice. *Crime and Justice* 16:99–180. [MBP]

Roberts, J. V. & Hough, M., eds. (2002) *Changing attitudes to punishment*. Willan. [MBP]

Roberts, J. V. & Stalans, L. J. (2004) Restorative sentencing: Exploring the views of the public. *Social Justice Research* 17(3):315–34. [MBP]

Roberts, S. C. (2012) *Applied evolutionary psychology*. Oxford University Press. [SCR]

Robinson, P., Kurzban, R. & Jones, O. D. (2007) The origins of shared intuitions of justice. *Vanderbilt Law Review* 60:1633–88. [MBP]

Rolls, E. T. (1990) A theory of emotion, and its application to understand the neural basis of emotion. *Cognition and Emotion* 4:161–90. [FA]

Rolls, E. T. (1995) A theory of emotion and consciousness, and its application to understanding the neural basis of emotion. In: *The cognitive neurosciences*, ed. M. S. Gazzaniga, pp. 1091–106. MIT Press. [FA]

Romero, T., Colmenares, F. & Aureli, F. (2009) Testing the function of reconciliation and third-party affiliation for aggressors in Hamadryas Baboons (*Papio hamadryas hamadryas*). *American Journal of Primatology* 71:60–69. [aMEMcC]

Ross, D. (2004) Meta-linguistic signalling for coordination amongst social agents. *Language Sciences* 26:621–42. [DR]

Ross, D. (2005) *Economic theory and cognitive science: Microexplanation*. MIT Press. [DR]

Ross, D. (2006) The economics and evolution of selves. *Journal of Cognitive Systems Research* 7:246–58. [DR]

Ross, D. (2007) H sapiens as ecologically special: What does language contribute? *Language Sciences* 29:710–31. [DR]

Ross, D. (2012) Special human vulnerability to low-cost collective punishment. *Behavioral and Brain Sciences* 35(1):37–38. [DR]

Rubin, D. C. (2005) A basic-systems approach to autobiographical memory. *Current Directions in Psychological Science* 14(2):79–83. [RF]

Rucker, D. D., Polifroni, M., Tetlock, P. E. & Scott, A. L. (2004) On the assignment of punishment: The impact of general-societal threat and the moderating role of severity. *Personality and Social Psychology Bulletin* 30:673–84. [FP]

Sahlins, M. (1972) *Stone age economics*. Aldine. [SB]

Sanfey, A. G. (2007) Social decision-making: Insights from game theory and neuroscience. *Science* 318(5850):598–602. [RF]

Schelling, T. (1960) *The strategy of conflict*. Harvard University Press. [arMEMcC]

Schmidt, K. M. (1993) Reputation and equilibrium characterization in repeated games of conflicting interests. *Econometrica* 61:325–52. [HG]

Schopler, J. & Compere, J. (1971) Effects of being kind or harsh to another on liking. *Journal of Personality and Social Psychology* 20(2):155–59. [ANS]

Schumann, K. & Ross, M. (2010) The benefits, costs, and paradox of revenge. *Social and Personality Psychology Compass* 4(12):1193–205. doi: 10.1111/j.1751-9004.2010.00322.x [SK, aMEMcC]

Schwartz, S. H. (2005) Robustness and fruitfulness of a theory of universals in individual human values. In: *Valores e comportamento nas organizações* [*Values and behavior in organizations*], ed. A. Tamayo & J. B. Porto, pp. 56–95. Vozes. [DL]

Schwarz, N. (2000) Emotions, cognition, and decision-making. *Emotion and Cognition* 14:433–40. [KO]

Schwarz, N. & Clore. G. L. (1983) Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology* 45:513–23. [KO]

Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P. & Johnson, M. (1997) Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature* 385(6613):254–57. [RY]

Scott-Philips, T. C. (2008) On the correct application of animal signalling theory to human communication. In: *Proceedings of the 7th International Conference on the Evolution of Language*, 7th edition, ed. A. D. M. Smith, K. Smith & R. Ferrer-i-Cancho, pp. 275–82. World Scientific. [rMEMcC]

Sell, A. (2011) The recalibrational theory and violent anger. *Aggression and Violent Behavior* 16(5):381–89. [aMEMcC, ANS]

Sell, A., Tooby, J. & Cosmides, L. (2009) Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences USA* 106(35):15073–78. [CH, arMEMcC, ANS]

Seyfarth, R. M. & Cheney, D. L. (2012) The evolutionary origin of friendship. *Annual Review of Psychology* 63:153–77. [FA]

Seyfarth, R. M. & Cheney, D. L. (in press) Knowledge of social relations. In: *Evolution of primate societies*, ed. J Mitani, J. Call, P. Kappeler, R. Palombit & J. Silk. University of Chicago Press. [FA]

Seymour, B., Singer, T. & Dolan, R. J. (2007) The neurobiology of punishment. *Nature Reviews Neuroscience* 8:300–11. [aMEMcC]

Sherif, M. (1966) *In common predicament: Social psychology of intergroup conflict and cooperation*. Houghton-Mifflin. [RJC]

Shinada, M., Yamagishi, T. & Ohmura, Y. (2004) False friends are worse than bitter enemies: "Altruistic" punishment of in-group members. *Evolution and Human Behavior* 25:379–93. [aMEMcC]

Silk, J. B. (1992) The patterning of intervention among male bonnet macaques: Reciprocity, revenge, and loyalty. *Current Anthropology* 33:318–25. [aMEMcC]

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

Silventoinen, K., Magnusson, P. K., Tynelius, P., Kaprio, J. & Rasmussen, F. (2008) Heritability of body size and muscle strength in young adulthood: A study of one million Swedish men. *Genetic Epidemiology* 32:341–49. [aMEMcC]

Simon, H. (1990) A mechanism for social selection and successful altruism. *Science* 250:1665–68. [HG]

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J. & Frith, C. D. (2006) Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439:466–69. [aMEMcC, SCR]

Smith, J. M. (1982) *Evolution and the theory of games*. Cambridge University Press. [rMEMcC]

Solomon, B. S., Bradshaw, C. P., Wright, J. & Cheng, T. L. (2008) Youth and parental attitudes toward fighting. *Journal of Interpersonal Violence* 23:544–60. [aMEMcC]

Speckhard, A. & Ahkmedova, K. (2006) The making of a martyr: Chechen suicide terrorism. *Studies in Conflict and Terrorism* 29:429–92. [aMEMcC]

Sperber, D. (1994) The modularity of thought and the epidemiology of representations. In: *Mapping the mind*, ed. L. Hirschfeld & S. Gelman, pp. 39–67. Cambridge University Press. [arMEMcC]

Sripada, C. S. & Stich, S. (2007) A framework for the psychology of norms. In: *The innate mind: Vol. 2, culture and cognition*, ed. P. Carruthers, S. Laurence & S. Stich, pp. 280–301. Oxford University Press. [CH]

Srivastava, J., Espinoza, F. & Fedorikhin, A. (2009) Coupling and decoupling of unfairness and anger in ultimatum bargaining. *Journal of Behavioral Decision Making* 22:475–89. [aMEMcC]

Steger, M. F., Hicks, B. M., Kashdan, T. B., Krueger, R. F. & Bouchard, T. J. (2007) Genetic and environmental influences on the positive traits of the values in action classification, and biometric covariance with normal personality. *Journal of Research in Personality* 41:524–39. [aMEMcC]

Stein, D. J. (1998) Psychiatric aspects of the Truth and Reconciliation Commission in South Africa. *The British Journal of Psychiatry; The Journal of Mental Science* 173:455–57. [DJS]

Stein, D. J. & Kaminer, D. (2006) Forgiveness and psychopathology: Psychobiological and evolutionary underpinnings. *CNS Spectrums* 11(2):87–89. [DJS]

Stein, D. J., Seedat, S., Kaminer, D., Moomal, H., Herman, A., Sonnega, J. & Williams, D. R. (2008) The impact of the Truth and Reconciliation Commission on psychological distress and forgiveness in South Africa. *Social Psychiatry and Psychiatric Epidemiology* 43(6):462–68. [DJS]

Stevens, J. R., Cushman, F. A. & Hauser, M. D. (2005) Evolving the psychological mechanisms for cooperation. *Annual Review of Ecology, Evolution, and Systematics* 36:499–518. [rMEMcC]

Stewart, E. A., Schreck, C. J. & Simons, R. L. (2006) 'I ain't gonna let no one disrespect me': Does the code of the street reduce or increase violent victimization among African American adolescents? *Journal of Research in Crime and Delinquency* 43:427–58. [aMEMcC]

Stillwell, A. M., Baumeister, R. F. & Del Priore, R. E. (2008) We're all victims here: Toward a psychology of revenge. *Basic and Applied Social Psychology* 30 (3):253–63. doi: 10.1080/01973530802375094 [SK]

Stouten, J., De Cremer, D. & Van Dijk, E. (2005) All is well that ends well, at least for proselfs: Emotional reactions to equality violation as a function of social value orientation. *European Journal of Social Psychology* 35:767–83. [DL]

Stouten, J., De Cremer, D. & Van Dijk, E. (2006) Violating equality in social dilemmas: Emotional and retributive reactions as a function of trust, attribution, and honesty. *Personality and Social Psychology Bulletin* 32(7):894–906. [DL]

Strayer, F. F. & Noel, J. M. (1986) The prosocial and antisocial functions of preschool aggression: An ethological study of triadic conflict among young children. In: *Altruism and aggression: Biological and social origins*, ed. C. Zahn-Waxler, E. M. Cummings & R. Iannotti, pp. 107–31. Cambridge University Press. [DP]

Struthers, C. W., Eaton, J., Santelli, A. G., Uchiyama, M. & Shirvani, N. (2008) The effects of attributions of intent and apology on forgiveness: When saying sorry might not help the story. *Journal of Experimental Social Psychology* 44:983–92. [aMEMcC]

Stuckless, N. & Goranson, R. (1992) The vengeance scale: Development of a measure of attitudes toward revenge. *Journal of Social Behavior and Personality* 7(1):25–42. [SK]

Summerfield, D. (2002) Effects of war: Moral knowledge, revenge, reconciliation, and medicalised concepts of "recovery." *British Medical Journal* 325:1105–107. [aMEMcC]

Sutton, R. M. & Winnard, E. J. (2007) Looking ahead through lenses of justice: The relevance of just-world beliefs to intentions and confidence in the future. *British Journal of Social Psychology* 46:649–66. [FP]

Svoboda, E., McKinnon, M. C. & Levine, B. (2006) The functional neuroanatomy of autobiographical memory: A meta-analysis. *Neuropsychologia* 44(12):2189–208. [RF]

Szymanska, E. J. (2011) *Retaliation versus vigilantism: Why do we choose to punish?* Doctoral dissertation, University of Pennsylvania. [rMEMcC]

Tabak, B. A., McCullough, M. E., Luna, L. R., Bono, G. & Berry, J. W. (2012) Conciliatory gestures facilitate forgiveness and feelings of friendship by making transgressors appear more agreeable. *Journal of Personality* 80:503–36. [aMEMcC]

Tannenbaum, D., Uhlmann, E.L. & Diermeier, D. (2011) Moral signals, public outrage, and immaterial harms. *Journal of Experimental Social Psychology* 47:1249–54. [ELU]

Taylor, D. M. & Moghaddam, F. M. (1987) *Theories of intergroup relations: International social psychological perspective.* Praeger. [RJC]

Tetlock, P. E. (2003) Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences* 7:320–24. [FP]

Tiger, L. (1969) *Men in groups*. Random House. [rMEMcC]

Tinbergen, N. (1963) On aims and methods of ethology. *Zeitschrift für Tierpsychologie* 20:410–33. [aMEMcC]

Todd, P. M. (2001) Fast and frugal heuristics for environmentally bounded minds. In: *Bounded rationality: The adaptive toolbox*, ed. G. Gigerenzer & R. Selten. pp. 51–70. MIT Press. [PB]

Tooby, J. & Cosmides, L. (1990a) On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality* 58:17–68. [arMEMcC]

Tooby, J. & Cosmides, L. (1990b) The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology* 11:375–424. [FA]

Tooby, J. & Cosmides, L. (1992) The psychological foundations of culture. In: *The adapted mind: Evolutionary psychology and the generation of culture*, ed. J. Barkow, L. Cosmides & J. Tooby, pp. 19–136. Oxford University Press. [arMEMcC]

Tooby, J. & Cosmides, L. (1996) Friendship and the Banker's Paradox: Other pathways to the evolution of adaptations for altruism. In: *Evolution of social behaviour patterns in primates and man. Proceedings of the British Academy, vol. 88*, ed. W. G. Runciman, J. M. Smith & R. I. M. Dunbar, pp. 119–43. Oxford University Press. [aMEMcC]

Tooby, J. & Cosmides, L. (2005) Conceptual foundations of evolutionary psychology. In: *The handbook of evolutionary psychology*, ed. D. M. Buss, pp. 5–67. Wiley. [CH]

Tooby, J. & Cosmides, L. (2008) The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In: *Handbook of emotions*, 3rd edition, ed. M. Lewis, J. M. Haviland-Jones & L. F. Barrett, pp. 114–37. Guilford Press. [arMEMcC]

Tooby, J., Cosmides, L. & Barrett, H. C. (2003) The second law of thermodynamics is the first law of psychology: Evolutionary developmental psychology and the theory of tandem, coordinated inheritances. *Psychological Bulletin* 129:858–65. [rMEMcC]

Tooby, J., Cosmides, L., Sell, A., Lieberman, D. & Sznycer, D. (2008) Internal regulatory variables and the design of human motivation: A computational and evolutionary approach. In: *Handbook of approach and avoidance motivation*, ed. A. Elliott, pp. 251–72. Erlbaum/Psychology Press. [RF, aMEMcC, ANS]

Tripp, T. M. & Bies, R. J. (2010) "Righteous" anger and revenge in the workplace: The fantasies, the feuds, the forgiveness. In: *International handbook of anger*, ed. M. Potegal, G. Stemmler & C. Spielberger, pp. 413–31. Springer. [MP]

Trivers, R. L. (1971) The evolution of reciprocal altruism. *Quarterly Review of Biology* 46(1):35–57. [AD, HG, aMEMcC]

Trivers, R. L. (1974) Parent–offspring conflict. *American Zoologist* 14:249–64. [aMEMcC]

Tsoudis, O. (2002) The influence of empathy in mock jury criminal cases: Adding to the affect control model. *Western Criminology Review* 4:55–67. [SCR]

Turkheimer, E. & Waldron, M. (2000) Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin* 126:78–108. [aMEMcC]

Turner, R. N., Hewstone, M., Voci, A. & Vonofakou, C. (2008) A test of the extended intergroup contact hypothesis: The mediating role of intergroup anxiety, perceived ingroup and outgroup norms, and the inclusion of the outgroup in the self. *Journal of Personality and Social Psychology* 95:843–60. [RJC]

Twenge, J. & Campbell, W. K. (2003) "Isn't it fun to get the respect that we're going to deserve?" Narcissism, social rejection, and aggression. *Personality and Social Psychology Bulletin* 29(2):261–72. [SK]

Twenge, J., Konrath, S., Foster, J., Campbell, W. K. & Bushman, B. (2008) Egos inflating over time: A cross-temporal meta-analysis of the Narcissistic Personality Inventory. *Journal of Personality* 76(4):875–902; discussion 903-828. doi: JOPY507 [pii] 10.1111/j.1467-6494.2008.00507.x [SK]

Uhlmann, E. L. & Zhu, L. (under review) Acts, persons, and intuitions: Person-centered cues and gut reactions to harmless transgressions. [ELU]

Uhlmann, E. L., Zhu, L. & Diermeier, D. (under review a) When actions speak volumes: Person-centered cues and moral outrage over racial bigotry. [ELU]

Uhlmann, E. L., Zhu, L. & Tannenbaum, D. (under review b) When it takes a bad person to do the right thing. [ELU]

Uniacke, S. (2000) Why is revenge wrong? *Journal of Value Inquiry* 34:61–69. [aMEMcC]

UNICEF Report (2010) *Children accused of witchcraft: An anthropological study of contemporary practices in Africa*. Report published by United Nations International Children's Emergency Fund, April 2010. [MP]

Van der Wal, R. C., Karremans, J. C. & Cillessen, A. H. N. (2012) A dual-process perspective on interpersonal forgiveness. Unpublished data. [JCK]

Van Lange, P. A. M., Ouwerkerk, J. & Tazelaar, M. (2002) How to overcome the detrimental effects of noise in social interaction: The benefits of generosity. *Journal of Personality and Social Psychology* 82:768–80. [aMEMcC]

van Schaik, C. P. & Aureli, F. (2000) The natural history of valuable relationships in primates. In: *Natural conflict resolution*, ed. F. Aureli & F. B. M. de Waal, pp. 307–33. University of California Press. [aMEMcC]

Verdier, R., ed. (1981a) *La Vengeance: Études d'ethnologies, d'histoire et de philosophie, vol. 1: Vengeance et pouvoir dans quelque sociétés extra-occidentales.* Éditions Cujas. [SB]

Verdier, R. (1981b) Le système vindicatoire. In: *La Vengeance: Études d'ethnologies, d'histoire et de philosophie, vol. 1: Vengeance et pouvoir dans quelque sociétés extra-occidentales*, ed. R Verdier, pp. 13–42. Éditions Cujas. [SB]

Verdier, R., ed. (1986) *La Vengeance: Études d'ethnologies, d'histoire et de philosophie, vol. 2: Vengeance et pouvoir dans quelque sociétés extra-occidentales.* Éditions Cujas. [SB]

Verdier, R. (2008) Vengeance, societies, and power in Amazonian societies. In: *Revenge in the cultures of lowland South America*, ed. S. Beckerman & P. Valentine, pp. 259–69. University of Florida Press. [SB]

Verdier, R. & Poly, J.-P., eds. (1984) *La Vengeance: Études d'ethnologies, d'histoire et de philosophie, vol. 3: Vengeance, pouvoirs et idéologies dans quelque civilizations de l'Antiquité.* Éditions Cujas. [SB]

Vidmar, N. & Miller, D. (1980) Social psychological processes underlying attitudes toward legal punishment. *Law and Society Review* 14:565–602. [KO]

Vossekuil, B., Fein, R. A., Reddy, M., Borum, R. & Modzeleski, W. (2002) *The final report and findings of the Safe School Initiative: Implications for the prevention of school attacks in the United States*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education, Safe and Drug-Free Schools Program, and U.S. Secret Service, National Threat Assessment Center. [aMEMcC]

Wagner, U., N'Diaye, K., Ethofer, T. & Vuilleumier, P. (2011) Guilt-specific processing in the prefrontal cortex. *Cerebral Cortex* 21(11):2461–70. [RF]

Wakefield, J. (1992) Disorder as harmful dysfunction: A conceptual critique of *DSM-III-R*'s definition of mental disorder. *Psychological Review* 99(2):232–47. [PB]

Walker, J. & Halloran, M. (2004) Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics* 7:235–47. [aMEMcC]

Waller, J. (2004) Our ancestral shadow: Hate and human nature in evolutionary psychology. Paper presented at the Conference to Establish the Field of Hate Studies, Spokane, WA, March 19, 2004. [Conference held at Gonzaga University and organized by the Institute for Action against Hate.] [ANS]

Watts, D. P. (1995) Post-conflict social events in wild mountain gorillas (*Mammalia, Hominoidea*) I. Social interactions between opponents. *Ethology* 100:139–57. [aMEMcC]

Watts, D. P. (2006) Conflict resolution in chimpanzees and the valuable-relationships hypothesis *International Journal of Primatology* 27:1337–64. [JCK, aMEMcC]

West, S. A., Griffin, A. S. & Gardner, A. (2007) Social semantics: Altruism, cooperation, mutualism, strong reciprocity, and group selection. *Journal of Evolutionary Biology* 20:415–32. [aMEMcC]

West, S. A., Mouden, C. E. & Gardner, A. (2011) Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior* 32:231–62. [arMEMcC]

Westcott, H. & Jones, D. (1999) Annotation: The abuse of disabled children. *Journal of Child Psychology and Psychiatry* 40(4):497–506. [ANS]

West-Eberhard, M. J. (1992) Adaptation: Current usages. In: *Keywords in evolutionary biology,* ed. E. F. Keller & E. A. Lloyd, pp. 12–18. Harvard University Press. [rMEMcC]

Wilkowski, B. M., Robinson, M. D. & Troop-Gordon, W. (2010) How does cognitive control reduce anger and aggression? The role of conflict monitoring and forgiveness processes. *Journal of Personality and Social Psychology* 98(5):830–40. doi: 10.1037/a0018962 [RF, JCK]

Williams, G. C. (1966) *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton University Press. [arMEMcC]

Wilson, D. S. (1998) Hunting, sharing, and multilevel selection: The tolerated theft model revisited. *Current Anthropology* 39:73–97. [HG]

Wilson, E. O. (1998) *Consilience: The unity of knowledge*. Knopf. [rMEMcC]

Wilson, M. & Daly, M. (1985) Competitiveness, risk taking, and violence: The young male syndrome. *Ethology and Sociobiology* 6:59–73. [rMEMcC]

Wilson, T. D., Lindsey, S. & Schooler, T. Y. (2000) A model of dual attitudes. *Psychological Review* 107(1):101–126. [RF]

Wilson, W. (1971) Reciprocation and other techniques for inducing cooperation. *Journal of Conflict Resolution* 15:167–96. [aMEMcC]

Worthington, E. L., Jr. (2005) More questions about forgiveness: Research agenda for 2005–2015. In: *Handbook of forgiveness*, ed. E. L. Worthington Jr., pp. 557–73. Routledge. [aMEMcC]

Worthington, E. L., Jr. & DiBlasio, F. (1990) Promoting mutual forgiveness within the fractured relationship. *Psychotherapy: Theory, Research, Practice, Training* 27:219–23. [aMEMcC]

Worthington., E. L., Jr., Witvliet, C. V. O., Pietrini, P. & Miller, A. J. (2007) Forgiveness, health, and well-being: A review of evidence for emotional versus decisional forgiveness, dispositional forgivingness, and reduced unforgiveness. *Journal of Behavioral Medicine* 30:291–302. [aMEMcC]

Wu, J. & Axelrod, R. (1995) How to cope with noise in the iterated prisoner's dilemma. *Journal of Conflict Resolution* 39:183–89. [aMEMcC]

Wu, J.-J., Zhang, B.-Y., Zhou, Z.-X., He, Q.-Q., Zheng, X.-D., Cressman, R. & Tao, Y. (2009) Costly punishment does not always increase cooperation. *Proceedings of the National Academy of Sciences USA* 106:17448–451. [aMEMcC]

Yamagishi, T. (1986) The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology* 51:110–16. [RJC, aMEMcC]

Yamagishi, T. (1988) Seriousness of social dilemmas and the provision of a sanctioning system. *Social Psychology Quarterly* 51:32–42. [aMEMcC]

Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S. & Cook, K. S. (2009) The private rejection of unfair offers and emotional commitment. *Proceedings of the National Academy of Sciences USA* 106:11520–23. [aMEMcC]

Yi, R., Buchhalter, A. R., Gatchalian, K. M. & Bickel, W. K. (2007) The relationship between temporal discounting and the prisoner's dilemma game in intranasal abusers of prescription opioids. *Drug and Alcohol Dependence* 87:94–97. [rMEMcC]

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M. & Damasio, A. (2010) Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron* 65(6):845–51. [RY]

Yu, R., Calder, A. J. & Mobbs, D. (submitted) Advantageous inequality leads to prosocial payoffs in humans. [RY]

Zajonc, R. B. (1968) Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology* 9:1–27. [KO]

Zechmeister, J. S., Garcia, S., Romero, C. & Vas, S. N. (2004) Don't apologize unless you mean it: A laboratory investigation of forgiveness and retaliation. *Journal of Social and Clinical Psychology* 23:532–64. [aMEMcC]

Zechmeister, J. S. & Romero, C. (2002) Victim and offender accounts of interpersonal conflict: Autobiographical narratives of forgiveness and unforgiveness. *Journal of Personality and Social Psychology* 82(4):675–86. [SK]

Zillikens, M. C., Yazdanpanah, M., Pardo, L. M., Rivadeneira, F., Aulchenko, Y. S., Oostra, B. A., Uitterlinden, A. G., Pols, H. A. P. & van Duijn, C. M. (2008) Sex-specific genetic effects influence variation in body composition. *Diabetologia* 51:2233–41. [aMEMcC]

Zillmann, D., Katcher, A. H. & Milavsky, B. (1972) Excitation transfer from physical exercise to subsequent aggressive behavior. *Journal of Experimental Social Psychology* 8:247–59. [VJK]

Zitek, E. M., Jordan, A. H., Monin, B. & Leach, F. R. (2010) Victim entitlement to behave selfishly. *Journal of Personality and Social Psychology* 98:245–55. [KO]