12

EVOLVED MECHANISMS FOR REVENGE AND FORGIVENESS

MICHAEL E. McCULLOUGH, ROBERT KURZBAN,
AND BENJAMIN A. TABAK

In this chapter, we describe our efforts to understand the functions of the cognitive systems that underlie humans' capacities for revenge and forgiveness. A better understanding of these concepts is not only scientifically interesting but socially important as well. In developed nations, the desire for revenge is cited as a causal factor in as many as 20% of homicides (Kubrin & Weitzer, 2003). Roughly 20% of the perpetrators of violent assault and criminal property damage in the United Kingdom cite the desire for revenge as a motive (Home Office, 2003), and 61% of U.S. school shootings between 1974 and June 2000 were vengeance-motivated (Vossekuil, Fein, Reddy, Borum, & Modzeleski, 2002). The desire for revenge also motivates people to enlist in terrorist organizations (Speckhard & Ahkmedova, 2006).

Perhaps because revenge is so closely linked to aggression and violence, it has been fashionable in Western thought since the Stoic (and, later, Christian) philosophers to view revenge as immoral, irrational, or both (Murphy, 2003). As if seeking to restate this dim view of revenge in modern, therapeutic terms,

We gratefully acknowledge the support of the National Institute of Mental Health Grant 5R01MH071258, a grant from the Fetzer Institute, and support from the Center for the Study of Law and Religion at Emory University to the first author.

social scientists in the past century have also promulgated the idea that the desire for revenge is indicative of psychological dysfunction (Murphy, 2003).

The "revenge as disease" conceit had a predictable effect on how forgiveness came to be studied empirically as well: If the desire for revenge is a disease, then perhaps forgiveness is the cure. Indeed, many of the earliest empirical studies on forgiveness were related to the use of interventions for promoting forgiveness in therapeutic settings (DiBlasio & Benda, 1991; Hebl & Enright, 1993). These treatments do promote forgiveness—and reduce psychological symptoms of anxiety and depression and boost self-esteem (Lundahl, Taylor, Stevenson, & Roberts, 2008)—but such facts do not even come close to proving that forgiveness is a "cure" for revenge.

Evolutionary research and scholarship cast considerable doubt on "disease" and "cure" conceits for conceptualizing the human capacities for revenge and forgiveness. In this chapter, we propose that revenge and forgiveness are the results of distinct psychological adaptations that evolved to solve specific adaptive problems. We posit that one or more revenge mechanisms evolved because of their efficacy in deterring interpersonal harms and that one or more forgiveness mechanisms evolved because of their efficacy in preserving valuable relationships despite those harms. Here, we attempt to define revenge and forgiveness in functional terms that will make them more amenable to evolutionary analysis (Williams, 1966), to describe the selection pressures that gave rise to them, and to outline what we think are the proximate causes and the computations involved when people make choices to forgive or to avenge a wrong.

EVOLUTION OF REVENGE

In the section that follows, we attempt to outline a functional approach to understanding revenge. In particular, we attempt to define revenge functionally rather than behaviorally and demonstrate the value of such a definitional approach. Second, we attempt to outline the evolutionary selection pressures that might have given rise to psychological mechanisms that produce revenge as a functional output.

Revenge: A Functional Definition

A great deal of research and writing has been devoted to revenge, and people have powerful intuitions about it. Still, we believe it is worthwhile to take a step back and reflect on the evolved function of putative revenge systems. At its heart, revenge solves a problem that is faced to varying degrees by many species: how to change other organisms' incentives to induce them to emit

benefits and refrain from imposing costs on oneself (see Chapter 3). To the extent that other organisms can learn that a target organism will retaliate (or conditionally benefit) as a function of their behavior, it is beneficial for the target organism to signal that it will do so. One (albeit imperfect) way to signal that one will retaliate if harmed (or benefit if helped) is to actually do so. If neural tissue is assembled that reliably motivates these sorts of contingent punishments and contingent rewards, it may boost lifetime reproductive fitness of its bearer and therefore evolve precisely because of these functions. Our analysis begins with this simple, but crucial, idea.

This notion contrasts with the way some philosophers have defined revenge, but we think some of the previous definitions create as many problems as they solve. For example, Govier (2002) wrote, "When we seek revenge, we seek satisfaction by attempting to harm the other (or associated persons) as a retaliatory measure" (p. 2). 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 claimed that "revenge is personal and noninstrumental: 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).

These definitions, because they are proximate and do not commit to any function, make no obvious predictions about the design features of the psychology of revenge. "Enjoyment" and other proximate explanations (see Govier's, 2002, definition) leave a promissory note for an ultimate explanation that must be paid. Why should revenge produce pleasure? For no organism except humans would we accept that an explanation for a behavior is that it brings enjoyment.

In short, functional thinking about cognition and behavior reminds us that there is no free lunch. Why would a species such as *Homo sapiens* engage in costly behavior such as revenge unless the mechanism that creates revenge was designed to produce benefits in the currency of fitness or is a by-product of a structure that does yield fitness payoffs (Andrews, Gangestad, & Matthews, 2002)? What could maintain revenge in humans' behavioral repertoire?

The definitions cited earlier also introduce problems related to intentionality. What does it mean that revenge involves an attempt to impose retaliatory harm on an aggressor? Does the word *attempt* imply a conscious and deliberate effort to make another individual suffer? Is consciousness necessary? Is deliberation necessary? Or can this striving to harm one's provoker be automatic and/or unconscious? And is this distinction critical in any case?

We think a functional definition of revenge can clarify some of these points. Biologists regularly define behavior functionally, as when 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). By designating a function, it becomes possible to search for evidence of the features—behavioral or physiological—that contribute to accomplishing the putative function.

In similar fashion, we define *revenge* functionally as behavior resulting from a mechanism designed to deter the imposition of costs on (or the withholding of benefits from) oneself or one's allies by the imposition of costs following a target's imposition of costs (or withholding of benefits), where costs and benefits are defined in terms of their effects on lifetime reproductive fitness. That is, revenge is a deterrence system designed to change others' incentives regarding the self and one's kin or allies (see Chapter 4 for a similar functional analysis of what the chapter authors call the *power behavioral system*). By imposing costs after harm (or withheld benefits), revenge signals that subsequent acts will be subject to the same contingent response, thereby altering others' incentives. We hypothesize that humans possess psychological adaptations designed specifically to produce revenge.

This functional definition has several important features. First, it replaces considerations of intentionality (e.g., whether the organism is deliberately or consciously attempting to do something) with considerations of design (e.g., what the system that motivates revenge was designed to do). Moreover, the concept of design makes powerful empirical commitments; adaptation is a strong claim (Williams, 1966), and to the extent that the psychological mechanisms do not show features that support a deterrence function, the hypothesis that humans possess an innate psychology of revenge is undermined.

Our definition of revenge incorporates as instances of revenge all retaliatory impositions of costs that are caused by a mechanism designed for this purpose, even acts that are not based on deliberation or awareness and even those that do not actually manage to deter anything (as when people behave aggressively toward a driver whom they perceive to have mistreated them on the road). Such a definition also permits a distinction between costs to the provoker that arise from design for that function versus costs that arise as a by-product. Harming a provoker is only revenge when the system that motivated the harmful behavior was crafted for that purpose. Avoiding a provoker to avert a second harm is not revenge, but avoiding a provoker to limit his or her access to benefits might be. Likewise, the phenomenon of *displaced aggression*, in which a victim of aggression proceeds later to harm a third party (Miller, Pedersen, Earleywine, & Pollock, 2003; see also Chapter 6, this volume), may not be revenge, even if the third party is a genetic relative or ally of the original aggressor. If displaced aggression of this nature is not produced by a system designed for deterrence but rather is produced by the psychological processes that Miller et al. (2003) implicated (e.g., residual arousal and postaggression rumination that lead to what are, essentially, cognitive errors)—that is to say,

if displaced aggression is a mere by-product of other psychological processes—then it is not revenge. As a side note, to us it is an open question whether some instances of triggered displaced aggression might actually reflect the operation of a revenge system. What we wish to point out here is that this "triggered displaced aggression as revenge" hypothesis—though it might be wrong—would likely never have been generated solely by relying on the standard, nonfunctional framework that researchers commonly use to understand displaced aggression.

Selection Pressures That Gave Rise to Mechanisms for Revenge

In an influential review article, Clutton-Brock and Parker (1995) noted that retaliation (which they called *punishment*) is common among nonhuman animals (for a more recent example, see Jensen, Call, & Tomasello, 2007). They speculated that retaliation yields fitness gains by reducing the probability that the targets of retaliation will repeat their injurious actions against the retaliator in the future. Consistent with Clutton-Brock and Parker's analysis, we hypothesize that natural selection gave rise to one or more comparable deterrence systems in humans. In this sense, the adaptive consequences of revenge come not from what revenge causes per se but from what it prevents. For illustrative purposes here, we distinguish among three types of deterrence. The first two, direct and indirect deterrence of aggression, involve deterring the imposition of costs. The third involves deterring the withholding of benefits.

Direct Deterrence

By *direct deterrence*, we mean that revenge discourages aggressors from harming the avenger a second time. The logic of direct deterrence is straightforward: If a potential aggressor must make a decision in which he or she can 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 can change the potential aggressor's incentives so that the expected value of the cost-imposing action on the potential victim is negative. Revenge can accomplish this transformation of expected value by conveying to an aggressor that the retaliatory infliction of fitness costs will exceed the potential benefits to be gained by aggressing against the potential victim a second time (see Chapter 3). Nevertheless, direct deterrence gives rise to strategic complications. For example, although revenge at Time 1 might predict revenge at Time 2, there is nothing that forces this to be true. An organism could be, for example, intermittently vengeful. This leads to well-known problems of signaling that one's vengeful dispositions are stable over time (Frank, 1988).

Experimental evidence in support of revenge's effectiveness as a direct deterrent comes from experiments involving economic games such as the sequential and iterated prisoner's dilemma (Axelrod, 1984). 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 (Clark & Sefton, 2001, Table 6), an observation that holds not only in the United States (Hayashi, Ostrom, Walker, & Yamagishi, 1999). However, because noncooperation and punishment are the same in the prisoner's dilemma, such findings must be interpreted with care.

In the *iterated* prisoner's dilemma, subjects play multiple rounds of the 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—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" or variants of it tend to elicit cooperation from experimental subjects (e.g., Wilson, 1971), hinting at their effectiveness in deterring defection.

Moreover, in an analysis of data from five different laboratory studies of dyadic negotiation in which partners played 250 consecutive trials during which they could either punish, reward, or withhold reward (and punishment) from each other, Molm (1997) found that the frequency with which retaliatory punishment was used (i.e., the infliction of punishment after one's negotiation partner had previously punished the actor) was positively associated with the frequency with which partners rewarded each other. Likewise, the use of punishment following nonreward (i.e., the withholding of benefits) was associated with higher rates of rewarding. These findings suggest that retaliatory infliction of punishments in response to punishments and the withholding of rewards creates a relational climate in which the exchange of reward is more frequent. In contrast, Molm reported that the frequency with which dyads punished *noncontingently* (i.e., 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 nonreward that it promotes cooperation.

In some situations, one can benefit from revenge's efficacy as a deterrent simply by possessing the ability to retaliate against one's interaction partners; it is not always necessary to retaliate directly. Work in behavioral economics

also illustrates this basic point. Consider the difference in play in the dictator game (DG) as opposed to the ultimatum game (UG). In both games, some amount of money, say $10, is to be divided between two people. In the DG, one person unilaterally decides how to split the money. In the UG, one person, the "proposer," proposes a split, and the other person, the "responder," can either accept that split or reject it, in which case both players receive nothing. Rejection in the UG is revenge; the cost imposed is the amount that the proposer allocated to him- or herself. It is not surprising that typical proposals in the UG (roughly 40% of the stake), in which revenge is possible, are larger than in the DG (roughly 20% of the stake), in which revenge is not possible (Forsythe, Horowitz, Savin, & Sefton, 1994).

Social psychology experiments also show how the prospect of suffering revenge can deter aggressors from harming the prospective avenger. In one study (Diamond, 1977), undergraduate men wrote an essay, which a confederate then derogated. Participants were then brought back to the laboratory 24 hours later and were given the opportunity to give 10 (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 the shocks themselves. People who believed that they could harm their insulting evaluators without the threat of retaliation gave stronger shocks to the evaluators.

The lessons of empirical studies on the direct deterrent effects of punishment are not always straightforward, however. For example, Fehr and List (2004) used the trust game, a two-step dyadic game in which an "investor" first entrusts a sum of money to a second person called the "trustee"; the money is then multiplied by some constant (often tripled) by the researchers. In the second step, the trustee is given the opportunity to return some amount of money to the investor. Fehr and List permitted investors to indicate a minimum amount of money they required from their trustees in return. If trustees failed to return that minimum amount, that amount was automatically deducted from the trustees' payoffs. Return transfers were highest when this punishment option was available but left unused by investors. Nonetheless, the majority of investors used the punishment option. Houser et al. (2008), using a similar design, found that the threat of punishment reduced the fraction of money trustees returned to investors even if the threat of punishment was applied as a result of a random process rather than as a decision on the part of the investor. Likewise, revenge in some experimental settings increases, rather than deters, noncooperation. Dreber, Rand, Fudenberg, and Nowak (2008) found that using punishment against noncooperators reduced players' gains in an iterated prisoner's dilemma, possibly because punishment in this experiment imposed large costs: The cost of punishment and the size of the damage it inflicts clearly influence revenge's deterrent effects.

Deterring Third Parties

Mechanisms for revenge may also have been naturally selected for their efficacy in deterring would-be aggressors by virtue of revenge's ability to signal the avenger's aggressive potential. Reputation is important for understanding how third-party deterrence might work. Ancestral humans lived in small, close-knit groups (Boehm, 2008) without the benefit of institutions for protecting individual rights, so a readiness to retaliate against interpersonal harms might have been an important component of people's social reputations. Researchers have documented the importance of defense of honor (i.e., more or less, the perceived ability to defend one's interests with violent force when necessary) and the revenge that it stimulates as a major cause of violence among individuals from many societies (for a review, see McCullough, 2008; see also Chapter 10, this volume).

Consistent with the idea that revenge is enacted partly out of reputational concerns, laboratory studies show that the psychological mechanisms that cause revenge are sensitive to the presence of third parties. Victims retaliate more strongly against their provokers 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 (Brown, 1968; Kim, Smith, & Brigham, 1998). Also, when two men get into an argument, the mere presence of a third person doubles the likelihood that the argument will escalate to a violent encounter (Felson, 1982).

Deterring the Withholding of Benefits

Finally, we think mechanisms for revenge might have been naturally selected because of their efficacy in changing others' behavior to increase the delivery of benefits (as opposed to only reducing harm). Public goods games are useful for illustrating how revenge can deter the withholding of benefits. In these games, a few (often four to six) participants receive an initial endowment of money and are instructed to choose how to split that endowment between two different pools. One pool is private, and participants simply keep any money they place in it. The other pool is shared; money placed into this pool is multiplied by some amount greater than one, and the resultant total is subsequently divided evenly among all the players in the group. Money maximizers keep everything in their private pools; aggregate group wealth is maximized when everyone contributes to the public pool. These games are social dilemmas because they create a tension between individual and group outcomes. (The fact that they involve groups rather than dyads is incidental.)

Yamagishi (1986) had subjects play public goods games in groups of four, repeated over 12 trials. He varied whether participants could punish other members of the group and varied the price of punishment, that is, the cost one had to spend to reduce another player's payoff by one unit. Players used the sanctioning system when it was provided, and, in its presence, players contributed greater amounts to the public good. However, these results do not distinguish the proximate motive, that is, whether sanctioning is instrumental (i.e., the result of a motive to increase one's benefits through the use of incentives) or vengeful (i.e., the result of a motive to impose costs on individuals who had an opportunity to deliver benefits but chose not to do so).

Fehr and Gächter's (2002) results help to clarify the proximate motive for punishment in this context. Fehr and Gächter ran a similar game with a few modifications, the most important of which was that players changed groups from round to round, so punishment could not be used to induce group members who were uncooperative in round r to benefit the subject in round $r+1$. Nevertheless, their results were similar to Yamagishi's (1986): Participants sanctioned uncooperative group members, and group members cooperated more when the punishment option was available (see also Anderson & Putterman, 2006). These results imply the operation of the revenge system, given that instrumental motives were ruled out. Fehr and Gächter would not agree. They coined a new term, *altruistic punishment*, to describe their findings.

Carpenter and Matthews (in press) conducted an experiment that contained an important control condition that helps to identify the limits of any altruism that might be present in so-called altruistic punishment. They ran noniterated public goods games and varied whether participants could punish members of their own groups or members of other people's groups. In the key treatment, the "one-way TPP" (third-party punishment) condition, almost no one punished. The fact that one-way third-party punishment was so minimal when directed toward noncooperators in groups to which the subjects themselves did not belong strongly suggests that without the possibility of revenge, people tend not to punish.

EVOLUTION OF FORGIVENESS

In the section to follow, we attempt to outline the basics of a functional approach to understanding forgiveness. As we did in outlining the basics of a functional approach to revenge, we offer a functional definition of forgiveness and illustrate its advantages over previous definitional approaches. We also outline the evolutionary selection pressures that might have given rise to psychological mechanisms that produce forgiveness as a functional output.

Forgiveness: A Functional Definition

Natural selection gave rise to one or more psychological systems that produce revenge, we posit, by virtue of the fitness payoffs associated with direct deterrence, third-party deterrence, and, possibly, deterrence of benefit-withholding. However, avengers trade off the potential benefits lost by virtue of any damage that revenge does to relations with the harm doer, and they incur the (probabilistic) costs associated with any counterrevenge that might ensue as the result of their revenge. We therefore presume that the revenge system is designed to adjust its operation in response to the potential costs and benefits associated with revenge in any particular instance. When the costs of revenge are too high relative to its expected deterrence benefits, an organism might pursue an alternative course of behavior—forgiveness being one of the more likely ones.

Over the past decade, the first author's research group has defined *forgiveness* as a set of motivational changes whereby an organism becomes (a) decreasingly motivated to retaliate against an offending relationship partner; (b) decreasingly motivated to avoid the offender; and (c) increasingly motivated by good will for, and a desire to reconcile with, the offender, despite the offender's harmful actions (McCullough, 2008; McCullough, Worthington, & Rachal, 1997). Here, we refine this definition by adding a functional addendum: that one or more "forgiveness systems" produce these motivational changes because of their efficacy during evolution in promoting the restoration of beneficial relationships in the aftermath of interpersonal harms.

This newly "functionalized" definition of forgiveness permits all of the important conceptual distinctions that other theorists (e.g., Enright & Coyle, 1998; Worthington, 2005) consider important (e.g., that forgiveness is different from forgetting an offense, denying its reality, condoning it, or attempting to minimize its significance), and it enables a tighter conceptual link between forgiveness and reconciliation than has previously been recognized. Many theorists have been careful to distinguish forgiveness from *reconciliation*, with the latter concept indicating a restoration of the relationship between offender and victim (Worthington, 2005). In light of the functional definition of forgiveness that we propose, it might be possible to forgive a harm doer (i.e., to experience motivational changes by which one becomes less vengeful, less avoidant, and more benevolently disposed toward him or her) without reconciling (i.e., restoring the relationship). Nevertheless, we reason that modern humans are capable of forgiving because ancestral humans who deployed this strategy enjoyed the fitness benefits that came from restoring potentially valuable relationships.

Nevertheless, forgiveness, like revenge, involves costs. Forgiveness prepares a victim to reenter constructive relations with a harm doer based on

the prospect of capturing benefits from that relationship, but forgiveness entails foregoing revenge and its deterrent effects. Forgiveness, therefore, involves a loss of gains from changing the harm doer's incentives, potentially inviting recidivism (e.g., see Gordon, Burton, & Porter, 2004) and attacks from those who see the opportunity to exploit the forgiver. In short, forgiveness undermines the function of the revenge system by undermining deterrence. Thus, a forgiveness system, like a revenge system, should be sensitive to costs and benefits, and these costs and benefits should have shaped the suite of proximate social–psychological factors that turn the system on and off.

Forgiveness: Selection Pressures

As noted earlier, we hypothesize that putative forgiveness systems evolved in response to selection pressures for restoring relationships that, on average, would have boosted lifetime reproductive fitness, a quality that researchers have called *relationship value* (de Waal, 2000). The role of relationship value in determining animals' propensity to forgive and/or reconcile after conflict has been demonstrated in many simulations of the evolution of cooperation among dyads and networks of individuals (e.g., Axelrod, 1984; Hruschka & Henrich, 2006). Similar findings (Koski, Koops, & Sterck, 2007; Watts, 2006) have emerged from behavioral studies of many mammalian species' postconflict conciliatory behaviors. It is in relationships in which substantial potential fitness gains are possible (e.g., kin, mates, allies, exchange partners) that forgiveness and/or reconciliation appear to be most common in nonhuman animals.

The benefits to lifetime reproductive fitness differ by relationship type. They might entail, of course, inclusive fitness benefits (Hamilton, 1964). After all, imposing costs on one's close genetic relatives directly impairs one's own inclusive fitness. Also, kin are most likely, all else being equal, to be the source of direct and reciprocal benefits for reasons associated with kin altruism. Therefore, one might expect forgiveness to be more likely in the context of kin relationships, with closer relatives being more easily forgiven than more distant ones.

Social organisms will also undergo selection pressure for forgiveness in the context of cooperation between nonrelatives when repeated encounters are likely (Axelrod, 1984; Trivers, 1971). Individuals who could forgive in such contexts would acquire two fitness benefits. First, forgiving isolated transgressions would have inhibited the *echo effect* (Axelrod, 1984), whereby individuals who are cooperatively disposed nevertheless become locked in costly cycles of retaliation when initial unintended defections occur due to noise. Second, individuals who can forgive their reciprocal altruism partners following defections would have been able to preserve their access to benefits that their partners would have been able to provide them and would have

spared themselves the costs associated with establishing new relationships with new individuals whose social dispositions would be unknown (Hruschka & Henrich, 2006). On average, it may simply be less costly to forgive some number of defections from a well-established relationship partner than to retaliate, or to withdraw from the relationship, following an isolated defection.

Indeed, in computer simulations of the evolution of reciprocal altruism—especially when the possibility of noise is assumed—evolutionarily stable strategies tend to be more forgiving than tit for tat, which responds to defection with defection and to cooperation with cooperation (Frean, 1994; Hauert & Schuster, 1998; Wu & Axelrod, 1995). This is especially true when one models reciprocal altruism as occurring largely among small networks of individuals (e.g., friendship groups, individuals within small living groups) who focus their cooperative efforts on other individuals within the network and limit their cooperation with individuals outside of the network (Levine & Kurzban, 2006). Under such circumstances, agents are expected to forgive up to 80% of other network members' defections (Hruschka & Henrich, 2006).

Other types of relationships generate still other types of benefits that redound to lifetime reproductive fitness. The benefits that might accrue from forgiving a mate are different from the benefits that might accrue by forgiving a friend, which in turn are different from the benefits that a forgiver might receive by forgiving an ally. Because the fitness-enhancing properties come in different currencies, the psychological systems that produce forgiveness are likely set up to identify the types of benefits that a particular type of relationship is likely to confer (and to weight them appropriately with respect to the probability of capturing those benefits, the time horizon at which they will be realized, etc.) and then weigh those benefits against the deterrent value of revenge, which the organism would trade off if it chose to forgive instead of seeking revenge (McCullough, Luna, Berry, Tabak, & Bono, in press).

CHOOSING FORGIVENESS OVER VENGEANCE: PROXIMATE CAUSATION

If, as we hypothesize, forgiveness systems are sensitive to tradeoffs associated with sacrificing the deterrence benefits of revenge for the relationship-restoration benefits of forgiveness, then such systems should be acutely sensitive to variables that influence the value of each option. These variables include, but are not necessarily limited to, characteristics of the offender, the transgression itself, and cues that predict the probabilities of future attacks and/or the potential future value of the restored relationship. In other words, we predict that forgiveness is generated by systems designed to compute and

compare the cost of forgone revenge and the benefits that are expected to accrue from a restored relationship.

Value of Deterrence

The value of revenge diminishes to the extent that it does not change behavior that would otherwise occur. In the limiting case, suppose that after an offense, the transgressor could persuasively signal that he or she would never—or could never—again inflict costs. In such a case, revenge would yield no benefit (except through third-party deterrence).

Information relevant to inferring intent can come from various sources. For instance, a transgressor's apology, expression of sympathy for a victim's suffering, and declaration of his or her intention to behave better in the future could indicate a low likelihood of trying to harm the victim in the future (McCullough et al., 1997). Verbal declarations such as these are susceptible to strategic manipulation, of course. Nonverbal displays such as blushing, which facilitate forgiveness after some transgressions (de Jong, Peters, & de Cremer, 2003), also contain information about changed intent and a transgressor's eagerness to distance himself or herself from a transgression, and their reliability may come from their unfakeability (Frank, 1988).

Other situational features might reduce the perceived deterrent value of revenge by convincing a victim that the transgressor's harmful actions were unintentional in the first place. People more readily forgive transgressors whose behavior was unintentional, unavoidable, or committed without awareness of the potential negative consequences (McCullough et al., in press). Also, it is unnecessary to engage in deterrence when additional transgressions are impossible. When the aggressor's capacity for violence is removed, for instance, vengeance yields little additional deterrent value. In some ethnographic accounts, reconciliation rituals involve the surrender of weapons (e.g., Boehm, 1987), which seems well suited to conveying an unwillingness to commit future aggressive acts. Trust may be a key psychological process by which the aforementioned factors that cue benevolent intentions lower the likelihood of revenge and raise the likelihood of forgiveness (Kurzban, 2003). People more readily forgive people whom they trust (Hewstone, Cairns, Voci, Hamberger, & Niens, 2006) and people who are reputed to be trustworthy (Vasalou, Hopfensitz, & Pitt, 2008) despite their recent bad behavior.

Expected Value of the Relationship

Against the costs of forgone revenge is the expected value of future benefits in a relationship in which intentions are positive rather than negative. 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. Because of the well-known principles of kin selection, close relatives are likely to be a source of benefits, and thus, we expect that cues of kinship will facilitate forgiveness, just as they evidently facilitate the restraint of vengeance (Lieberman & Linke, 2007).

In similar fashion, those with whom one has a close history of association, shared interests, and many opportunities for mutually beneficial transactions are good candidates for forgiveness because of the possibility of continued gains. Indeed, priming people with the names of close others (e.g., via subliminal presentation) leads to increased judgments of forgiveness, increased accessibility of the concept of forgiveness, and reduced deliberation about whether forgiveness is an appropriate course of action (Karremans & Aarts, 2007). Karremans and Aarts's (2007) results complement those from several previous studies showing that people are more inclined to forgive individuals with whom they feel close and committed (Finkel, Rusbult, Kumashiro, & Hannon, 2002; McCullough et al., 1998; see also Chapter 2, this volume). We would argue that the reason for these associations of closeness and/or commitment with forgiveness is that relationship closeness and commitment act as cues of relationship value in many types of relationships. We think the importance of relationship value can also explain why people tend to want some form of compensation prior to forgiving (Boehm, 1987; Bottom, Gibson, Daniels, & Murnighan, 2002): Compensation may serve as a cue of (among other things) a transgressor's ability or willingness to be a valuable relationship partner in the future.

In computations of expected value, empathy may play a special role. Empathy for transgressors, which is a sympathy- or pity-like response to the plight of another person, appears to be a reliable facilitator of forgiveness (McCullough et al., 1997), perhaps as a result of empathy's long phylogenetic history as a motivator of care for valuable relationship partners (Preston & de Waal, 2002). Whether the empathy–forgiveness link is part of the design of the forgiveness system or merely an incidental effect that empathy can exert within the existing forgiveness system, however, is currently difficult to know.

SUMMARY

The desire for revenge and the ability to forgive seem to be universal psychological endowments of humans (Boehm, 2008; Daly & Wilson, 1988; McCullough, 2008). Species-typical traits call out for explanations in terms of the mind's evolved mental structures, either as direct products or as by-products of what those structures were designed to do (Andrews et al., 2002). Here, we have taken an adaptationist stance and posited that revenge and forgiveness result from computational mechanisms designed to produce them. Once one has moved into a functional framework, we think it becomes easier to see what should qualify as revenge and forgiveness—and what should not—and what the important questions are if one wants to understand what revenge and forgiveness are really all about. By outlining the selection pressures that likely gave rise to humans' penchants for revenge and forgiveness, we have also tried here to identify the types of information that the structures that produce revenge and forgiveness should be designed to process. We hope that introducing this sort of thinking can help investigators prioritize their research efforts in the future.

REFERENCES

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. doi:10.1016/j.geb.2004.08.007

Andrews, P. W., Gangestad, S. W., & Matthews, D. (2002). Adaptationism—How to carry out an exaptationist program. *Behavioral and Brain Sciences, 25*, 489–504. doi:10.1017/S0140525X02000092

Axelrod, R. (1984). *The evolution of cooperation*. New York, NY: Basic Books.

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*, 139–147. doi:10.1037/h0044242

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

Boehm, C. (2008). Purposive social selection and the evolution of human altruism. *Cross-Cultural Research, 42*, 319–352. doi:10.1177/1069397108320422

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. doi:10.1287/orsc.13.5.497.7816

Brown, B. R. (1968). The effects of need to maintain face on interpersonal bargaining. *Journal of Experimental Social Psychology, 4*, 107–122. doi:10.1016/0022-1031(68)90053-X

Carpenter, J. P., & Matthews, P. H. (in press). Norm enforcement: Anger, indignation, or reciprocity? *Journal of the European Economic Association*.

Clark, K., & Sefton, M. (2001). The sequential prisoner's dilemma: Evidence on reciprocation. *The Economic Journal, 111*, 51–68. doi:10.1111/1468-0297.00588

Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature, 373*, 209–216. doi:10.1038/373209a0

Daly, M., & Wilson, M. (1988). *Homicide*. New York, NY: Aldine de Gruyter.

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*, 225–249. doi:10.1023/A:1025059631708

de Waal, F. B. M. (2000, July). Primates: A natural heritage of conflict resolution. *Science, 289*, 586–590. doi:10.1126/science.289.5479.586

Diamond, S. R. (1977). The effect of fear on the aggressive responses of anger aroused and revenge motivated subjects. *Journal of Psychology, 95*, 185–188.

DiBlasio, F., & Benda, B. B. (1991). Practitioners, religion and the use of forgiveness in the clinical setting. *Journal of Psychology and Christianity, 10*, 166–172.

Dreber, A., Rand, D. G., Fudenberg, D., & Nowak, M. A. (2008, March 20). Winners don't punish. *Nature, 452*, 348–351. doi:10.1038/nature06723

Elster, J. (1990). Norms of revenge. *Ethics, 100*, 862–885. doi:10.1086/293238

Enright, R. D., & Coyle, C. T. (1998). Researching the process model of forgiveness within psychological interventions. In E. L. Worthington (Ed.), *Dimensions of forgiveness: Psychological research and theological perspectives* (pp. 139–161). Philadelphia, PA: Templeton Foundation Press.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137–140. doi:10.1038/415137a

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*, 743–771. doi:10.2139/ssrn.364480

Felson, R. B. (1982). Impression management and the escalation of aggression and violence. *Social Psychology Quarterly, 45*, 245–254. doi:10.2307/3033920

Finkel, E. J., Rusbult, C. E., Kumashiro, M., & Hannon, P. A. (2002). Dealing with a betrayal in close relationships: Does commitment promote forgiveness? *Journal of Personality and Social Psychology, 82*, 956–974. doi:10.1037/0022-3514.82.6.956

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior, 6*, 347–369. doi:10.1006/game.1994.1021

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

Frean, M. R. (1994). The prisoner's dilemma without synchrony. *Proceedings of the Royal Society of London B, 257*, 75–79. doi:10.1098/rspb.1994.0096

Gordon, K. C., Burton, S., & Porter, L. (2004). Predicting the intentions of women in domestic violence shelters to return to partners: Does forgiveness play a role? *Journal of Family Psychology, 18*, 331–338. doi:10.1037/0893-3200.18.2.331

Govier, T. (2002). *Forgiveness and revenge*. New York, NY: Routledge.

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

Hauert, C., & Schuster, H. G. (1998). Extending the iterated prisoner's dilemma without synchrony. *Journal of Theoretical Biology, 192*, 155–166. doi:10.1006/jtbi.1997.0590

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. doi:10.1177/104346399011001002

Hebl, J., & Enright, R. D. (1993). Forgiveness as a psychotherapeutic goal with elderly females. *Psychotherapy, 30*, 658–667. doi:10.1037/0033-3204.30.4.658

Hewstone, M., Cairns, E., Voci, A., Hamberger, J., & Niens, U. (2006). Intergroup contact, forgiveness, and experience of "The Troubles" in Northern Ireland. *Journal of Social Issues, 62*, 99–120. doi:10.1111/j.1540-4560.2006.00441.x

Home Office. Research, Development and Statistics Directorate. Offending Surveys and Research, National Centre for Social Research and BMRB. (2003). *Social Research, Offending, Crime and Justice Survey, 2003* [computer file] (Report No. SN: 5248, 3rd ed.). Retrieved from http://www.data-archive.ac.uk/doc/5248%5Cmrdoc%5CUKDA%5CUKDA_Study_5248_Information.htm

Houser, D., Xiao, E., McCabe, K., & Smith, V. (2008). When punishment fails: Research on sanctions, intentions and noncooperation. *Games and Economic Behavior, 62*, 509–532.

Hruschka, D. J., & Henrich, J. (2006). Friendship, cliquishness, and the emergence of cooperation. *Journal of Theoretical Biology, 239*, 1–15. doi:10.1016/j.jtbi.2005.07.006

Jensen, K., Call, J., & Tomasello, M. (2007). Chimpanzees are vengeful but not spiteful. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 13046–13050. doi:10.1073/pnas.0705555104

Karremans, J. C., & Aarts, H. (2007). The role of automaticity in determining the inclination to forgive close others. *Journal of Experimental Social Psychology, 43*, 902–917. doi:10.1016/j.jesp.2006.10.012

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*, 353–361. doi:10.1177/0146167298244002

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*, 158–172. doi:10.1002/ajp.20338

Kubrin, C. E., & Weitzer, R. (2003). Retaliatory homicide: Concentrated disadvantage and neighborhood culture. *Social Problems, 50*, 157–180. doi:10.1525/sp.2003.50.2.157

Kurzban, R. (2003). Trust, reciprocity, and gains from association: Interdisciplinary lessons from experimental research. In E. Ostrom & J. Walker (Eds.), *Biological foundations of reciprocity* (pp. 105–127). New York, NY: Sage.

Levine, S. S., & Kurzban, R. (2006). Explaining clustering in social networks: Towards an evolutionary theory of cascading benefits. *Managerial and Decision Economics, 27*, 173–187. doi:10.1002/mde.1291

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

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*, 465–478. doi:10.1177/1049731507313979

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford, England: Oxford.

McCullough, M. E. (2008). *Beyond revenge: The evolution of the forgiveness instinct*. San Francisco, CA: Jossey-Bass.

McCullough, M. E., Luna, L. R., Berry, J. W., Tabak, B. A., & Bono, G. (in press). On the form of forgiving: Modeling the time-forgiveness relationship and testing the valuable relationships hypothesis. *Emotion*.

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships. II: Theoretical elaboration and measurement. *Journal of Personality and Social Psychology, 75*, 1586–1603. doi:10.1037/0022-3514.75.6.1586

McCullough, M. E., Worthington, E. L., & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology, 73*, 321–336. doi:10.1037/0022-3514.73.2.321

Miller, N., Pedersen, W. C., Earleywine, M., & Pollock, V. E. (2003). A theoretical model of triggered displaced aggression. *Personality and Social Psychology Review, 7*, 75–97. doi:10.1207/S15327957PSPR0701_5

Molm, L. D. (1997). *Coercive power in social exchange*. New York, NY: Cambridge University Press.

Murphy, J. G. (2003). *Getting even: Forgiveness and its limits*. New York, NY: Oxford.

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*, 1–72.

Rapoport, A., & Chammah, A. M. (1965). *The prisoner's dilemma*. Ann Arbor: University of Michigan Press.

Speckhard, A., & Ahkmedova, K. (2006). The making of a martyr: Chechen suicide terrorism. *Studies in Conflict and Terrorism, 29*, 429–492. doi:10.1080/10576100600698550

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology, 46*, 35–57. doi:10.1086/406755

Uniacke, S. (2000). Why is revenge wrong? *The Journal of Value Inquiry, 34*, 61–69. doi:10.1023/A:1004778229751

Vasalou, A., Hopfensitz, A., & Pitt, J. V. (2008). In praise of forgiveness: Ways for repairing trust breakdowns in one-off online interactions. *International Journal of Human-Computer Studies, 66*, 466–480. doi:10.1016/j.ijhcs.2008.02.001

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.

Watts, D. (2006). Conflict resolution in chimpanzees and the valuable-relationships hypothesis. *International Journal of Primatology, 27*, 1337–1364. doi:10.1007/s10764-006-9081-9

Williams, G. C. (1966). *Adaptation and natural selection. A critique of some current evolutionary thought*. Princeton, NJ: Princeton University Press.

Wilson, W. (1971). Reciprocation and other techniques for inducing cooperation. *The Journal of Conflict Resolution, 15*, 167–195. doi:10.1177/002200277101500205

Worthington, E. L. (2005). More questions about forgiveness: Research agenda for 2005–2015. In E. L. Worthington (Ed.), *Handbook of forgiveness* (pp. 557–573). New York, NY: Routledge.

Wu, J., & Axelrod, R. (1995). How to cope with noise in the iterated prisoner's dilemma. *The Journal of Conflict Resolution, 39*, 183–189. doi:10.1177/0022002795039001008

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*, 110–116. doi:10.1037/0022-3514.51.1.110