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