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.

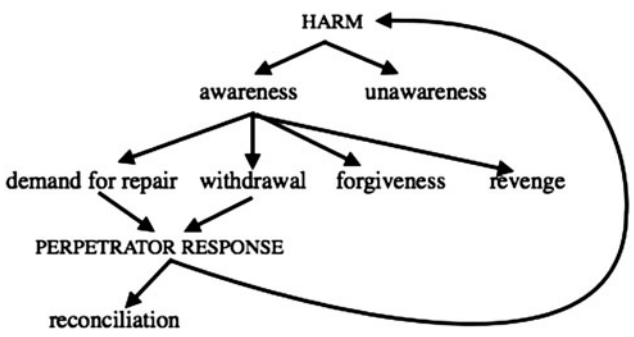

Figure 1 (Johnson-Freyd & Freyd). Responses to interrelational harm.

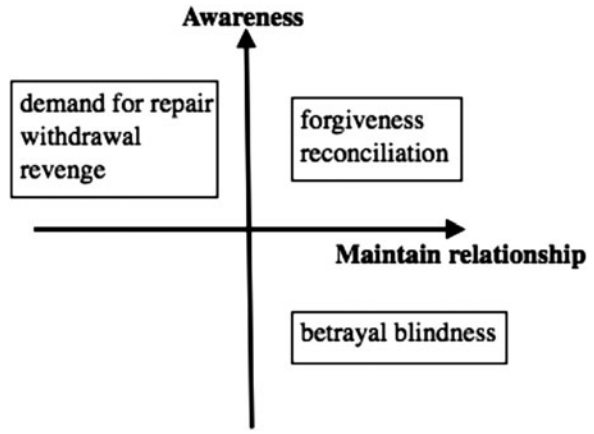

Figure 2 (Johnson-Freyd & Freyd). Proposed dimensions of responses to interrelational harm.

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.