Mutualists don't look back in anger

E. Glen Weyl[1], Megan E. Frederickson[1,2], Douglas W. Yu[3,4,‡], Naomi E. Pierce[1,5]

[1]*Society of Fellows, Harvard University, 78 Mount Auburn Street, Cambridge, Massachusetts 02138, USA.* [2]*Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Ontario M5S 3G5, Canada.* [3]*State Key Laboratory of Genetic Resources and Evolution, Ecology, Conservation, and Environment Center (ECEC), Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China.* [4]*School of Biological Sciences, University of East Anglia, Norwich, Norfolk NR47TJ, UK.* [5]*Department of Organismic & Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA.* [‡]*Corresponding author: Douglas W. Yu (dougwyu@gmail.com)*

Although mutualisms are common in all ecological communities and have played key roles in the diversification of life, our current understanding of the evolution of cooperation[1-3] applies mostly to social behavior within a species, while mutualism theory has lagged behind. A major question is whether mutualisms persist because hosts have evolved costly punishment of cheaters. Here, we use the economic theory of employment contracts to formulate and distinguish between two mechanisms that have been proposed to prevent cheating in host-symbiont mutualisms, Partner Fidelity Feedback[1,4,5] (PFF) and Host Sanctions[1,6] (HS). Under PFF, positive feedback between host fitness and symbiont fitness is sufficient to prevent cheating; in contrast, HS requires costly punishment to maintain mutualism. A coevolutionary model of mutualism finds that HS are unlikely to evolve *de novo,* and published data on legume-rhizobia and yucca-moth mutualisms are consistent with PFF and not HS. Thus, in systems considered to be textbook cases of HS, we find poor support for the theory that hosts have evolved to 'look back in anger' and punish cheating symbionts; instead, we show that even horizontally transmitted mutualisms can be stabilized via PFF. PFF theory may place previously underappreciated constraints on the evolution of mutualism, and explain why punishment is far from ubiquitous in nature.

PFF occurs when the benefits provided by a donor individual to a recipient individual feed back to the donor[1,4]. The harder the donor works to assist the recipient, the better off the recipient is and the more benefits it, in turn, provides back to the donor. Analogously, if a donor harms the recipient, the harm also feeds back to the donor. PFF is possible only when partners associate long enough that the short-term costs of helping can be recuperated by the helper. Under PFF, natural selection favors mutualists rather than cheaters because an individual that fails to cooperate reduces its own fitness; no further punishment is necessary. By contrast, HS posits that PFF is not sufficient to negate the incentive to cheat, and thus mutualism will persist only if hosts evolve to punish cheaters[6] (Box 1). Although both concepts (PFF and HS) have been discussed in some form in the literature at least since Trivers' seminal paper in 1971[7], it was Bull and Rice[8] who coined the term 'Partner Fidelity' in a review paper in 1991, which they distinguished from Partner Choice (PC). Today, PC is often used interchangeably with HS, but Bull and Rice originally defined PC to mean interactions in which individuals "differentially reward cooperative vs. un-cooperative partners *in advance of any possible exploitation*" [italics added], whereas both PFF and HS are differential rewards or punishments implemented *after* exploitation is possible. Here, we will argue that a failure to clearly define the differences between PFF and HS has led to their conflation, with the result that experiments demonstrating the punishment of cheating in mutualisms[9-14] have been generally accepted as evidence for HS[1,4,11-15], whereas PFF is the more likely explanation.

Consider a symbiotic *agent* (A) that may either take a cooperative *action* a^* that is in its host's interest or cheat by taking one of two alternative actions, a_1 or a_2. The host, the *principal* (P), then observes some (imperfect) *signal* s of the action taken by A. Some signals, S_1 and S_2, are *smoking guns* in that they indicate that A has cheated by performing a_1 or a_2, respectively. Based on the signal, P decides whether to punish A and if so, by how much, π. A natural strategy for P is to evolve a *punishment policy* $\pi^*(s)$ that maximizes its fitness, $U^P(s,\pi)$, for each signal s. A's fitness following any particular action increases with the short-term benefits of cheating, $u^s(a)$, plus the feedback benefits that A derives from P's fitness, scaled by f, and decreases with the punishment π that A receives. The expected fitness of any action $U^A(a;\pi)$ thus depends on P’s punishment policy. If, for simplicity, we let these have independent influences on A, then

$$U^A(a;\pi) = u^s(a) + fU^P[s;\pi] - \pi \tag{1}$$

This model allows us to differentiate PFF and HS once we have made an important, if common, assumption: subject to the constraints imposed by the theories, both A and P maximize their fitness. In such an evolutionarily stable equilibrium, P would, in the absence of active retribution, always follow policy π^*. Therefore, *PFF is the theory that P uses* π^* *and*

$$U^A(a^*;\pi^*) \geq U^A(a_1;\pi^*),\ U^A(a_2;\pi^*) \tag{2}$$

That is, “natural” feedback is enough to sustain observed cooperation.

On the other hand, to say that a_1 *is disciplined by HS* is to insist that an additional punishment mechanism must evolve:

$$U^A(a^*;\pi^*) < U^A(a_1;\pi^*) \tag{3}$$

Evolutionary stability under HS then requires that the punishment necessary to maintain cooperation be administered in the least costly manner, a scheme derived by Holmström[16] in his theory of moral hazard.

Under what conditions would we expect HS to evolve? In Supplementary Material 1, we extend a one-sided HS model[6,15] to a two-sided, coevolutionary model and find two evolutionarily stable strategies (ESS): (1) HS are absent and (2) HS are maintained, but only when the symbiont population is a mix of cheaters and cooperators and only for a limited set of parameter values. For the system to reach (2), it would need to somehow escape the first ESS of no sanctions and no cooperative agents. Alternatively, HS could evolve if the host induces evolution in the symbiont, something that is possible, but more likely to occur in highly specific situations with strong spatial structure and/or hosts with generation times much longer than those of the symbiont (Supplementary Material 1). If the most likely starting point for a symbiosis is a largely homogeneous population of initially non-mutualistic symbionts and non-investing, non-sanctioning hosts, then costly punishment by hosts is unlikely to evolve *de novo*. An exogenous input of cheaters or biased mutation can maintain HS, but cannot explain the origin of HS[17].

In the absence of HS, PFF (Eqn. 2) can maintain mutualism if *f* takes a value that is initially high because of host and symbiont pre-adaptations (Eqn. 1). By definition, an initially high value of *f* does not evolve in response to the existence of the symbiotic relationship, but is determined exogenously by the life-cycle concordance of the partners[1] (Eqn. 1). Although it is usually posited that PFF arises from spatial structuring or vertical transmission, *f* can also take high values due to widespread pre-adaptations in plants that limit, withdraw, or even abscise investments in shoots, flowers, and roots upon physical damage or deficits of pollen or other resources (e.g. Refs. 11,18-20,21 and included references). If a sufficient amount of symbiont fitness is also reduced after reducing investment in a plant part, the conditions for PFF are met (Eqn. 2), and symbionts are selected to protect, pollinate, or fertilize in order to maintain plant investment.

Using Holmström's[16] solution for optimal punishment, we design two experimental tests to distinguish PFF and HS (Supplementary Material 2) that require less information and are easier to implement than is characterizing the full fitness functions of A and P. **Test 1:** Suppose there are two signals, s and s', that give the same prospective incentive for punishment ($\pi^*(s) = \pi^*(s')$), but $s \in S_1$ indicates cheating, while $s' \notin S_1 \cup S_2$ does not. PFF predicts equal punishment of s and s' while HS requires that s be punished more ($\pi_{HS}(s) > \pi_{HS}(s')$). **Test 2:** Suppose that a_2 provides at least as much benefit to A as a_1, that, on average, P is at least as fit after a_2 as after a_1, but that a_2 is punished less under π^*. PFF predicts less punishment of a_2 than of a_1, while HS predicts that a_2 receives at least as much punishment as a_1; otherwise a_1 is being punished more than necessary to deter cheating.

We can use these tests to infer whether PFF or HS maintains cooperation in two well known mutualisms in which hosts are thought to punish cheater symbionts. The first is the mutualism between legume plants and rhizobia bacteria, which nodulate plant roots and expend energy to ‘fix’ atmospheric nitrogen (N_2) into a form (NH_4^+) that the plant can use to build proteins[22]. In return, the plant provides carbohydrates to the rhizobia. Some rhizobia genotypes cheat by fixing little or no N_2 and accumulating energy reserves for reproduction instead[22]. Some authors have argued[1,4,6,10,13,15] that HS disciplines cheating in rhizobia. When soybean plant nodules were exposed to an atmosphere in which N_2 had been replaced by Ar, thereby preventing nitrogen fixation, nodules grew to smaller sizes and supported fewer rhizobia than did controls exposed to the normal N_2-rich atmosphere. Kiers et al.[10] assumed that plants cannot, on a nodule-by-nodule basis, measure the concentration of N_2 and thus cannot distinguish cheating by rhizobia from a lack of N_2. If this assumption is true, then **Test 1** can be used to distinguish between PFF and HS.

Let s be the signal of low N_2 fixation in the argon experiment, indicating cheating under the above assumption. Let s' be the signal from a second experiment that reduces the value of rhizobia to the host plant by the same amount as the argon experiment but could not possibly indicate cheating. Such a signal is provided by experiments in which the nitrogen requirements of host plants are fully met via fertilizers. PFF predicts equal reduction in rhizobia populations, whereas HS predicts greater punishment in the argon treatment. If we compare what happens when soybean root systems are 1) exposed to an $Ar{:}O_2$ atmosphere that permits little N_2 fixation or 2) grown in high levels of nitrates, almost the same, reduced rhizobial densities are reported[23], which supports PFF. We emphasize two caveats. In the published experiment[23], we have assumed that the level of added nitrates fully met the soybean’s short-term nitrogen requirements. This might not be strictly true. Second, the fact that legumes are often infected by multiple strains of rhizobia remains a problem, because our model assumes only one type of agent.

The second mutualism is that between yucca plants and yucca moths, which actively pollinate yucca flowers but also lay eggs so that their larvae can consume the developing seeds. Selection should favor moths that lay more eggs, at the cost of seeds, which results in a conflict of interest between plant and moth. Pellmyr and Huth[9] documented that *Yucca filamentosa* selectively abscises flowers with high egg loads (and low pollen loads). Again, using **Test 1**, let s be the signal of high egg load. Let s' be the signal from a second experiment[20], in which pins are used to mimic the puncture damage to ovules caused by moth oviposition, while omitting the eggs themselves. PFF predicts equal levels of floral abscission across the two experiments, but HS predicts greater punishment when eggs are present. Marr and Pellmyr[20] found that mechanical damage alone was sufficient to trigger levels of floral abscission consistent with naturally observed levels[9] and that the plant did not appear to react to the experimental application of moth eggs. Consistent with PFF, selective abortion of flowers is thought to be a “symplesiomorphy within the Agavaceae, and thus a preadaptation within the yuccas.”[19] Ovule damage is also correlated with probability of selective abortion in an independently evolved pollination mutualism between *Glochidion* trees and *Epicephala* moths[24].

The yucca system also provides an example of **Test 2**. *Yucca kanabensis* is associated with two moth species, one of which oviposits shallowly in the carpel wall and does not damage ovules[25]. ‘Shallow’ species execute a_2 because they achieve equal or greater cheating benefits (on average, slightly more surviving larvae per locule) relative to ‘deep’ species, which execute a_1 by also damaging ovules during oviposition. Furthermore a_2 results in an equally fit principal (“equivalent numbers of mature, viable seeds”[25]). Consistent with PFF, *Y. kanabensis* only selectively abscises ‘deep’ species[25] whose damage to ovules directly degrades flower value.

In summary, our analysis suggests that well known examples of mutualism, previously used as examples of HS, are probably more consistent with PFF. We conclude that PFF is the more general mechanism for maintaining costly mutualistic behavior in symbionts and can be selected for even under situations of horizontal transmission. Costly punishment in mutualisms is less likely to evolve *de novo* (HS) than are the natural consequences of the presence of the appropriate pre-adaptations (or spatial structures), PFF[11,19]. Thus PFF might place previously under-appreciated constraints on the evolution of mutualism. More generally, biological mutualisms and the economics of mechanism design form a surprisingly natural marriage of data and theory that provides novel directions for empirical studies and the possibility of further insights into the evolution of cooperation.

We thank Drew Fudenberg, Jerry Green, Marco Archetti, Moshe Hoffman, Joy Bergelson, and seminar participants at Harvard and at the University of Chicago for comments. We received funding from NSF SES-0750480, from the TECT initiative of ESF/EUROCORES, and from the Yunnan provincial government (20080A001) and the Chinese Academy of Sciences (0902281081). EGW and MEF were supported by the Harvard Society of Fellows.

Supplementary Material 1. ESS models for the evolution of punishment.

Supplementary Material 2. Differentiating PFF from HS.

Box 1. PFF vs. HS. One way to contrast PFF and HS is by analogy to the criminal justice system. PFF is akin to the theory that incarceration should bring about incapacitation and/or rehabilitation, thus preventing any prospective (short-term future) harms the criminal may also cause society, and only incidentally punishing misbehavior. Each case is justified individually. By contrast, HS are analogous to theories that emphasize the value of punishment as a public deterrent to criminal activity. According to Becker's economic theory of crime and punishment[26], the least expensive means of deterrence are punishments that are infrequently administered, widely publicized, and severe, such as corporal punishment or torture. The prospect of suffering similar punishment is transmitted culturally to the wider population and deters misbehavior by others. In a biological context, HS is the theory that the symbiont population is 'evolutionarily improved' by host punishment. See also analysis in Supplementary Material 1.

An example is recent reaction to illegal downloading of copyrighted media, such as films and music. As in HS, copyright owners have tried to deter such violations by prosecuting individuals who can be shown in court to have shared files illegally and publicizing the cases. However, recent years have shown that socially acceptable levels of punishment (typically, small fines) are too low to be an effective deterrent. Alternatively, simple non-retributive policies may help reduce violations at low social cost. Illegal downloads consume substantial bandwidth, so some have argued that establishing an automatic system of graduated charges for bandwidth consumption would significantly reduce such violations. Such policies are efficient even if illegal downloading is not taking place, as they help ensure efficient use of scarce bandwidth. Thus, as with PFF, natural incentives without any retributive motive may be sufficient to deter antisocial behavior.

1 Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. The evolution of cooperation. *Q Rev Biol* 79, 135-160 (2004).

2 Nowak, M. A. Five rules for the evolution of cooperation. *Science* 314, 1560 – 1563 (2006).

3 West, S. A., Griffin, A. S. & Gardner, A. Evolutionary explanations for cooperation. *Current Biology* 17, R661-R672 (2007).

4 Foster, K. R. & Wenseleers, T. A general model for the evolution of mutualisms. *Journal of Evolutionary Biology* 19, 1283-1293 (2006).

5 Frederickson, M. E. Conflict over reproduction in an ant-plant symbiosis: why Allomerus octoarticulatus ants sterilize Cordia nodosa trees. *Am Nat* 173, 675-681, doi:10.1086/597608 (2009).

6 West, S. A., Kiers, E. T., Simms, E. L. & Denison, R. F. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc R Soc Lond B* 269, 685-694 (2002).

7 Trivers, R. L. The evolution of reciprocal altruism. *Q Rev Biol* 46, 35-57 (1971).

8 Bull, J. J. & Rice, W. R. Distinguishing mechanisms for the evolution of co-operation. *Journal of Theoretical Biology* 149, 63-74 (1991).

9 Pellmyr, O. & Huth, C. J. Evolutionary stability of mutualism between yuccas and yucca moths. *Nature* 372, 257-260 (1994).

10 Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume-rhizobium mutualism. *Nature* 425, 78-81 (2003).

11 Edwards, D. P., Hassall, M., Sutherland, W. J. & Yu, D. W. Selection for protection in an ant-plant mutualism: host sanctions, host modularity, and the principal-agent game. *Proc R Soc Lond B* 273, 595-602, doi:10.1098/rspb.2005.3273 (2006).

12 Bever, J. D., Richardson, S. C., Lawrence, B. M., Holmes, J. & Watson, M. Preferential allocation to beneficial symbiont with spatial structure maintains mycorrhizal mutualism. *Ecol Lett* 12, 13-21, doi:10.1111/j.1461-0248.2008.01254.x (2009).

13 Simms, E. L. *et al.* An empirical test of partner choice mechanisms in a wild legume-rhizobium interaction. *Proc R Soc Lond B* 273, 77-81 (2006).

14 Jandér, K. C. & Herre, E. A. Host sanctions and pollinator cheating in the fig tree - fig wasp mutualism. *Proc R Soc Lond B* (in the press).

15 West, S. A., Kiers, E. T., Pen, I. & Denison, R. F. Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? *J Evol Biol* 15, 830-837 (2002).

16 Holmström, B. Moral hazard and observability. *Bell. J. Econ* 10, 74-91 (1979).

17 Foster, K. R. & Kokko, H. Cheating can stabilize cooperation in mutualisms. *Proc R Soc Lond B* 273, 2233–2239, doi:10.1098/rspb.2006.3571 (2006).

18 Jandér, K. C. & Herre, E. A. Host sanctions and pollinator cheating in the fig tree - fig wasp mutualism. *Proc R Soc Lond B*, doi:10.1098/rspb.2009.2157 (2010).

19 Pellmyr, O. & Huth, C. J. Differential abortion in the yucca: Reply. *Nature* 376, 558 (1995).

20 Marr, D. L. & Pellmyr, O. Effect of pollinator-inflicted ovule damage on floral abscission in the yucca-yucca moth mutualism: the role of mechanical and chemical factors. *Oecologia* 136, 236-243 (2003).

21 Hodge, A. The plastic plant: root responses to heterogeneous supplies of nutrients. *New Phytologist* 162, 9-24 (2004).

22 Kiers, E. T. & Denison, R. F. Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Ann Rev Ecol Evol Syst* 39, 215-236 (2008).

23 **Kiers, E. T., Rousseau, R. A. & Denison, R. F. Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. *Evol Ecol Res* 8, 1077-1086 (2006).**

24 **Goto, R., Okamoto, T., Kiers, E. T., Kawakita, A. & Kato, M. Selective flower abortion maintains moth cooperation in a newly discovered pollination mutualism. *Ecol Lett* 13, 321–329 (2010).**

25 **Addicott, J. F. & Bao, T. Limiting the costs of mutualism: multiple modes of interaction between yuccas and yucca moths. *Proc R Soc Lond B* 266, 197-202 (1999).**

26 **Becker, G. S. Crime and punishment: an economic approach. *J. Pol. Econ.* 76, 169-217 (1968).**