

# The Coevolution of Punishment and Prosociality Among Learning Agents

Fiery Cushman (cushman@wjh.harvard.edu)

Department of Psychology, 1484 William James Hall, 33 Kirkland St.,  
Cambridge, MA 02138 USA

Owen Macindoe (owenm@mit.edu)

Computer Science and Artificial Intelligence Laboratory, Building 32-G585, 32 Vassar Street  
Cambridge, MA 02139 USA

## Abstract

We explore the coevolution of punishment and prosociality in a population of learning agents. Across three models, we find that the capacity to learn from punishment can allow both punishment and prosocial behavior to evolve by natural selection. In order to model the effects of innate behavioral dispositions (such as prosociality) combined with the effects of learning (such as a response to contingent punishment), we adopt a Bayesian framework. Agents choose actions by considering their probable outcomes, calculated from an innate, heritable prior distribution and agents' experience of actual outcomes. We explore models in which an agent learns about the dispositions of each individual agent independently, as well as models in which an agent combines individual-level and group-level learning. Our results illustrate how the integration of Bayesian cognitive models into agent-based simulations of natural selection can reveal evolutionary dynamics in the optimal balance between innate knowledge and learning.

**Keywords:** costly punishment; hierarchical Bayesian models; prosociality; evolution; societal modeling.

## Introduction

Why isn't theft more common? To respect others' property is nice, but pilfering is more profitable. One family of explanations posits a key role for punishment (e.g. Boyd & Richerson, 1992). In brief, when people can learn from punishment, it pays to punish theft. And, when people punish theft, it pays not to steal. Thus, the coevolution of punishment and prosociality depends critically upon the manner in which perpetrators learn from punishment. Here, we compare the evolutionary dynamics that result under several different learning models, demonstrating sometimes dramatically different outcomes. These findings have implications for our understanding of the evolution of social behavior in particular, but also provide a case study of a more general problem: how can cognitive models of learned behavior and evolutionary models of innate behavior be integrated?

We model agents equipped with Bayesian learning mechanisms. One key advantage of a Bayesian approach is that it provides a natural means of combining heritable, innate dispositions with experience to determine an agent's behavior (see also Kirby, Dowman, & Griffiths, 2007). In our model, an agent decides whether or not to steal from a potential victim by considering the probability that it will be punished for theft, and then using this probability to calculate the expected value of theft (the antisocial choice) versus abstaining from theft (the prosocial choice). Agents' guesses about the probability of punishment are generated by a Bayesian updating

mechanism: their innate belief about the probability of punishment (i.e., a prior distribution over hypotheses) along with their experience of past punishment (i.e., data) is used to assess the probability of future punishment (i.e. the probability of a new datum given the posterior distribution over hypotheses). Thus, it should be clear how we can use prior probabilities to model innate prosociality versus antisociality. A prior expectation of punishment will translate into an innate prosocial tendency not to steal. Conversely, a prior expectation against punishment will translate into an innate antisocial tendency to steal. Critically, we allow heritable prior probabilities to mutate in a system of replicating agents—thereby changing innate contributions to prosocial vs. antisocial behavior—and explore the resultant evolutionary dynamics.

Modeling innate behavioral tendencies as prior probabilities allows for these tendencies to vary in strength and persistence—in biological terms, innate tendencies can vary in their degree of canalization versus plasticity. For instance, an agent could have an innate belief that it will not be punished that is very strong (highly canalized), in which case it will continue to steal in the face of consistently experienced punishment for theft. Or, an agent could have an innate belief that it will be punished that is very weak (highly plastic), in which case it will cease stealing after experiencing a small amount of consistent punishment. This feature of the Bayesian approach provides an important advantage over past attempts to model the coevolution of punishment and prosociality. Other studies have considered a restricted set of specialized behavioral responses to punishment: for instance, adopting pure prosocial behavior after a single instance of punishment, (e.g. Sigmund, Hauert, & Nowak, 2001; Brandt, Hauert, & Sigmund, 2003). This corresponds to a high degree of plasticity—but in the model we present, rather than stipulating a high degree of plasticity we explore the evolutionary dynamics that lead to plasticity versus canalization.

A second key advantage of a Bayesian approach is the ease with which so-called “hierarchical” learning can be modeled (Griffiths, Kemp, & Tenenbaum, 2008; Tenenbaum, Griffiths, & Kemp, 2006). For instance, in our model a potential thief might guess whether a victim is likely to punish by considering not only that specific victim's prior record of punishment, but also the prior record of punishment from other victims. Thus, inferences about individual-level behavior might be hierarchically embedded in inferences about group-level

behavior. Experimental evidence suggests that people do, indeed, adopt more prosocial behavior with novel group members when past group members have punished antisocial behavior (Fehr & Gächter, 2002). We directly compare learning models operating exclusively at the individual level with learning models that operate across the individual and group levels, demonstrating notably different evolutionary dynamics. Specifically, when everybody else punishes, and people predict that you will do what everybody else does, there is no need to pay the costs of punishment. Thus, a non-punitive strategy can invade a population dominated by costly punishment when inferences from group-level behavior to individual-level behavior are strong.

Throughout this paper we model an agent’s innate behavior in terms of its innate beliefs. We do not choose this approach because we believe that all innate behaviors are a consequence of innate beliefs, as such. Rather, we choose this approach because Bayes’ rule provides a framework in which innate and experienced factors can be formally combined, and in which innate factors can vary both in their consequences (i.e. degree of prosociality) and strength (i.e. degree of canalization). Thus, a formal model of beliefs can yield a useful functional model of behavior. We will consider both the benefits and drawbacks of this approach.

## Methods

We explored evolutionary dynamics for three different cognitive models through computer simulation. Each simulation involved a population size of 25 agents over 4000 generations. In every generation, each agent interacts with each other agent 50 times: 25 times in the “perpetrator” role (deciding whether or not to steal) and 25 times in the “victim” role (deciding whether or not to respond to theft with punishment).

Each time a perpetrator decides whether to steal, it has perfect knowledge of the value of the theft (always 1 fitness unit) as well as the value of some behavioral alternative to theft,  $V_{alt}$ , randomly chosen from  $(-2, 2)$  fitness units at the outset of each interaction. It also has perfect knowledge of the cost of being punished (always 2 fitness units), but must estimate the probability of being punished for theft,  $P(pun|theft)$ . It estimates this probability by one of three Bayesian methods, described below. It then calculates the expected value of theft,  $E_{theft} = 1 - 2P(pun|theft)$ . Finally, the agent chooses whether to steal by applying a softmax function with temperature parameter  $\lambda = 3$  to the expected value of theft and known value of the behavioral alternative (Daw & Doya, 2006):

$$P(theft) = \frac{e^{\lambda E_{theft}}}{e^{\lambda E_{theft}} + e^{\lambda V_{alt}}} \quad (1)$$

If the perpetrator chooses not to steal, no social interaction takes place and therefore the victim does not punish. If the perpetrator chooses to steal, the victim punishes with probability  $\theta$ . An agent’s  $\theta$  is fixed over the course of its lifetime, and inherited by that agent’s “children” subject to a 1% chance of mutation; mutant  $\theta$ s are drawn uniformly from the

interval  $(0, 1)$ . Punishment carries a cost of 0.5 fitness units to the punisher.

At the end of each generation all agents die and are replaced using a modified version of the pairwise comparison process for natural selection (Traulsen, Pacheco, & Nowak, 2007). In this process, each agent’s fitness,  $F_a$ , is compared with the fitness of another agent selected randomly,  $F_b$ , and the agent passes its heritable traits to the replacement with probability:

$$P(self\ replacement) = \frac{1}{1 + e^{F_b - F_a}} \quad (2)$$

Otherwise, the randomly selected other passes its heritable traits to the replacement.

### Model 1: Uniform prior

In the uniform prior model, perpetrators estimate the probability of being punished for theft  $P(pun|theft)$  for each victim independently, given its history of interactions with that victim. This is calculated from a posterior probability distribution over the interval of possible  $\theta$  values  $[0, 1]$  that victim might possess, given a uniform prior probability for all  $\theta$  values and a binomial distribution parameterized by the victim-specific history of punishment: the number of times  $N_{pun}$  that the agent has been punished by the victim for theft, and the total number of thefts  $N_{theft}$  from the victim. The posterior predictive distribution of punishment given theft—that is, the agent’s best guess as to whether its next theft will be punished by that individual—has a simple analytic solution (Griffiths et al., 2008):

$$P(pun|theft) = \frac{N_{pun} + 1}{N_{theft} + 2} \quad (3)$$

Thus, a perpetrator estimates  $P(pun|theft)$  at .5 in its first interaction with that agent, and will tend to make more accurate predictions over the course of subsequent interactions.

### Model 2: One-level

A simple extension of the uniform prior model allows the evolution of non-uniform prior probability distributions over  $\theta$ , so that individual perpetrators apply an inherited prior to each individual victim and calculate a victim-specific posterior distribution over  $\theta$  given that victim’s history of punitive responses. We define the prior distribution over  $\theta$  using a beta distribution parameterized  $Beta(\alpha, \beta)$ ,  $\alpha, \beta > 0$ . Here, again, the posterior predictive distribution (agent’s estimate  $P(pun|theft)$ ) has an analytic solution (Griffiths et al., 2008):

$$P(pun|theft) = \frac{(N_{pun} + \alpha)}{(N_{theft} + \alpha + \beta)} \quad (4)$$

The agent’s initial estimate  $P(pun|theft)$  is given by  $\frac{\alpha}{\alpha + \beta}$ . When this initial estimate is large, the agent exhibits an innate bias towards prosociality; when this initial estimate is small, the agent exhibits an innate bias towards antisociality. Additionally, large values of  $\alpha$  and  $\beta$  make an agent’s estimated  $P(pun|theft)$  resistant to experiential modification—that is, large values of  $\alpha$  and  $\beta$  dominate  $N_{pun}$  and  $N_{theft}$ . Thus,

canalization is captured by  $(\alpha + \beta) \gg 0$ , and plasticity by  $(\alpha + \beta) \approx 0$ .

In this one-level model, an agent’s  $\alpha$  and  $\beta$  parameters are heritable traits. During replacement,  $\alpha$  and  $\beta$  parameters are each subject to a 1% rate of mutation, with new values selected from exponential distributions with mean 2. In the first generation of each simulation, all individuals were initialized with parameters  $\alpha = 1, \beta = 1$ , that is, a uniform prior over  $\theta$ .

### Model 3: Hierarchical

Finally, we consider a hierarchical Bayesian learning model that combines inferences across the individual and group levels. In essence, a hierarchical model tests whether individuals’  $\theta$  values are consistent or inconsistent across the group, and makes inferences about individuals’ future behavior based on that decision. Thus, an agent’s prior history of punishment with one victim can exert an influence on its predictions about another victim’s punitive behavior. In order to perform this inference, each perpetrator calculates the posterior probability of several different probability distributions over  $\theta$  (hereafter, “hypotheses”). For instance, consider three particular hypotheses: (1): *Beta(100,1)*, (2): *Beta(1,100)* and (3): *Beta(1,0.01)*. If an agent consistently experiences punishment from all victims of theft, it will select (1) as the most likely hypothesis; conversely, if it never experiences punishment, it will select (2). However, if an agent experiences a mixed population where some victims consistently punish and others consistently do not, it will select (3) as the most likely hypothesis, because (3) allows the posterior probability of  $\theta$  to be strongly determined by the unique history of punishment with each individual victim.

In a hierarchical model, an individual’s behavior is characterized by high levels of prosociality when it favors hypotheses with values of  $\frac{\alpha}{\alpha+\beta} \approx 1$  and an individual’s behavior is characterized by strong “group bias” (i.e., tendency to infer from group behavior to individual behavior) when it favors hypotheses with values of  $(\alpha + \beta) \gg 0$ . Agents inherit an innate probability distribution over hypotheses, subject to mutation. If this probability distribution is even across all hypotheses, prosociality and group bias are highly plastic traits. As this innate probability distribution is increasingly skewed across hypotheses, either prosociality, group bias, or both become increasingly canalized.

In the hierarchical model we implemented, each agent considers 45 different probability distributions over  $\theta$ : a full crossing of nine levels of prosociality ( $\frac{\alpha}{\alpha+\beta} \in \{.1, .2, \dots, .9\}$ ) with five levels of group bias ( $\alpha + \beta \in \{0.2, 2, 20, 200, 2000\}$ ). An estimate of  $P(\text{pun}|\text{theft})$  is obtained by applying equation 4 to each of the 45 hypotheses, and then taking the average estimate weighted by the posterior probability of each hypotheses,  $P(h_i|\text{history})$ . For a given prior distribution over the 45 hypotheses  $P(h)$ , and a history of punitive responses, the posterior probability of each hypothesis can be calculated:

$$P(h_i|\text{history}) = \frac{P(\text{history}|h_i)P(h_i)}{P(\text{history})} \quad (5)$$

Where  $P(\text{history}|h_i)$  is the likelihood of each victim’s responses under a given hypothesis as calculated by equation 4, and  $P(\text{history})$  is the average  $P(\text{history}|h_i)$  for all 45 hypotheses, weighted by their prior probabilities.

We allowed replication and mutation in the prior probabilities assigned to the 45 hypotheses that each hierarchical learner considered. In order to simplify the evolution of prior probabilities over this hypothesis space, we calculated each agent’s prior probabilities with two heritable parameters: a “prosociality” parameter  $\psi$  and a “group bias” parameter  $\gamma$ . All 45 hypotheses were classified according to nine levels of  $\frac{\alpha}{\alpha+\beta}$  ranked from -5 to 5 ( $R_p$ ) and five levels of  $(\alpha + \beta)$  ranked from -2 to 2 ( $R_c$ ). The probabilistic weighting of each hypothesis was calculated:

$$S_{h_i} = e^{\psi R_{p_i}} e^{\gamma R_{c_i}} \quad (6)$$

Scores were then normalized to sum to 1 in order to derive the prior probability of each hypothesis. In the first generation of each simulation, all individuals were initialized with a uniform prior over the 45 hypotheses.

### Non-learning control models

As a control, we also tested three corresponding non-learning models: a uniform prior non-learning model, a one-level non-learning model, and a hierarchical non-learning model. Each of these models was constructed by eliminating the calculation of posterior probability distributions given a history of punishment. Thus, in the uniform-prior non-learning model, all agents always estimated the probability of punishment at exactly .5. In the one-level and hierarchical non-learning models, agents always estimated the probability of punishment according to their inherited prior distribution. The latter two cases test whether punishment and prosociality can co-evolve in the absence of learning.

## Results

### Uniform prior model

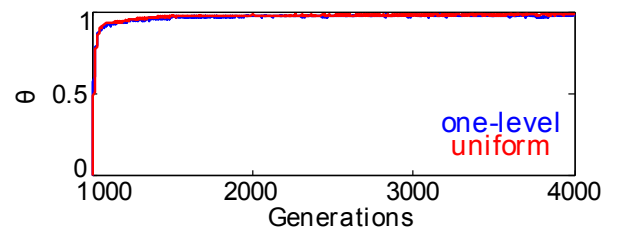


Figure 1: Average population  $\theta$  values for 50 simulations over generations 1000-4000 (following the introduction of  $\theta > 0$  as an evolvable strategy in generation 1000) for the uniform prior model and the one-level model.

Following the introduction of mutations in punishment strategies, population levels of punishment rapidly climbed towards the maximum level in the uniform prior model. That

is, once  $\theta$  was heritable and could mutate towards values  $> 0$ , population  $\theta$  values approached 1 within a few hundred generations (Figure 1). In the final, 4000th generation of each simulation, the average population  $\theta$  across 50 independent simulations was .98 (SE  $< .01$ ).

Throughout the results section, we describe “population parameters” (e.g., “population  $\theta$ ”) meaning the average value of the parameter over the 25 individuals comprising the population of a single simulation. Often, we discuss the “average population parameter” across 50 simulations, referring to independent simulations of 50 different populations, each comprising 25 individuals. That is, we assess whether multiple independent simulations are characterized by similar average values of the parameters of interest.

### One-level model

As in the uniform prior model, in the one-level model population levels of punishment climbed towards the maximum level ( $\theta \approx 1$ ) within a few hundred generations, and the strategy  $\theta \approx 1$  typically continued to dominate the population for the remainder of the simulation (Figure 1). In the final, 4000th generation of each simulation, the average population  $\theta$  across 50 independent simulations was .98 (SE  $< .01$ ).

Unlike the uniform prior model, the one-level model allowed prior probabilities of punishment to mutate and evolve. At the end of the initial 1000 generations of each simulation, during which  $\theta$  values were fixed at 0, the average population  $\frac{\alpha}{\alpha+\beta}$  across 50 distinct simulated populations was .05 (SE  $< .01$ ), indicating a behavioral bias towards antisocial behavior. The average population  $\alpha + \beta$  was 2.23 (SE .17), indicating low levels of canalization of this innate antisocial bias.

By the 2000th generation, however—following 1000 generations during which population  $\theta$  values climbed towards 1—this innate behavioral tendency had reversed. The average population  $\frac{\alpha}{\alpha+\beta}$  was .91 (SE .01), indicating a behavioral bias towards prosocial behavior. The average population  $\alpha + \beta$  was 3.2 (SE .25), indicating low levels of canalization of this innate prosocial bias. For the remaining 2000 generations of each simulation, population levels of prosociality and canalization remained largely stable.

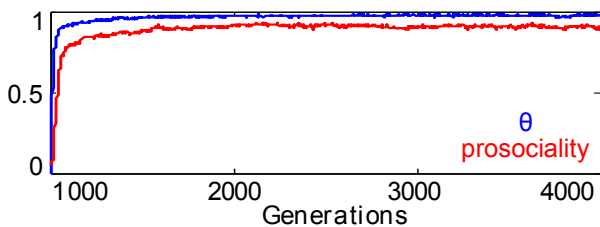


Figure 2: Average population levels of  $\frac{\alpha}{\alpha+\beta}$  and  $\theta$  for 50 simulations over generations 1000-4000 (following the introduction of  $\theta > 0$  as an evolvable strategy in generation 1000) for the one-level model.

As Figure 2 reflects, population values of  $\frac{\alpha}{\alpha+\beta}$  (innate estimate of others’ punishment levels) typically tracked popula-

tion values of  $\theta$  (innate punishment level) closely.

### Hierarchical model

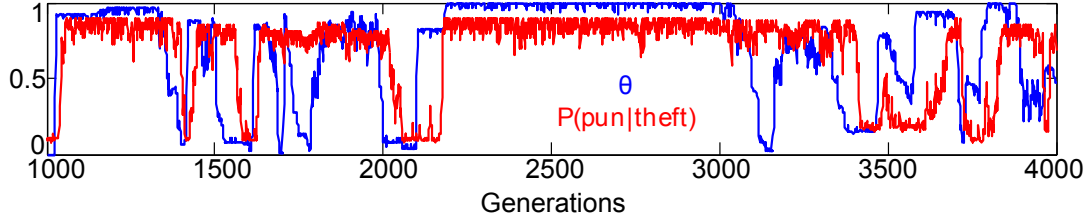
Unlike the one-level models, in which population levels of both punishment and prosociality climbed rapidly towards 1 (always punish theft, always expect theft to be punished) and then remained relatively constant, simulations using a hierarchical model typically exhibited large, synchronized fluctuations of punishment and prosociality. A representative example of a single simulation is plotted in Figure 3a. (Because averaging across distinct simulations masks these fluctuations, we have chosen to present graphical data from only a single simulation. We selected an example with slightly more fluctuations than typical, in order to provide several viewable instances for the reader. However, the degrees of synchronization exhibited in Figures 3a and 3b were chosen from the median).

Figure 3a exhibits the relationship between population levels of  $\theta$  and population levels of the prior  $P(\text{pun}|\text{theft})$ , which reflects an innate bias towards prosociality or antisociality. At the end of the initial 1000 generations of each simulation, during which  $\theta$  values were fixed at 0, the average population prior  $P(\text{pun}|\text{theft})$  across 50 simulated populations was .12 (SE  $< .01$ ), indicating a behavioral bias towards antisocial behavior. Over the course of the remaining 3000 simulations, population levels of  $\theta$  fluctuated between 0 and 1. And, the population prior  $P(\text{pun}|\text{theft})$  typically tracked these fluctuating  $\theta$ s with a slight delay. Thus, innate expectations of punishment tended to track actual punishment levels.

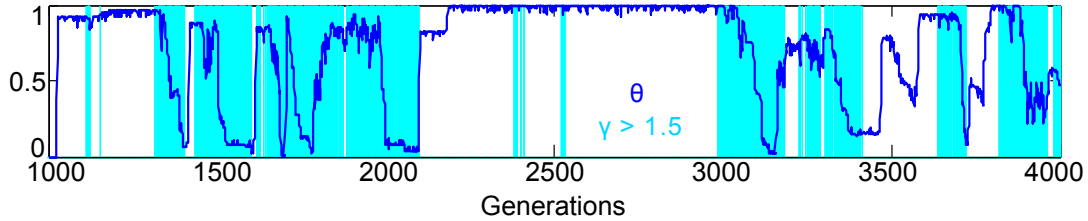
Figure 3b exhibits the relationship between population levels of “group bias” and population levels of  $\theta$ . The bright blue underlay indicates generations during where the heritable group bias parameter  $\gamma$  exceeded a value of 1.5. During these periods, individuals tended to draw strong inferences about the behavior of an individual based on their experience with others in the group. High levels of group bias were typically followed by a drop in population levels of  $\theta$ , while low levels of group bias were typically followed by an increase in population levels of  $\theta$ . Thus, punishment was favored most strongly when levels of group bias were low.

A possible explanation for the relationship between group bias and punishment, discussed further below, can be succinctly stated: when most people punish, and potential perpetrators infer that you’ll do what most people do, you can get away with not punishing and avoid its associated fitness costs.

If this explanation is correct then, given enough interactions, it should be possible for perpetrators to learn who does and who does not punish (overcoming a strong innate group bias). In order to test this prediction, we ran 10 independent simulations of the hierarchical model where each agent considers stealing from each other agent 100 times, rather than the standard 25 times used above. At the final, 4000th generation, average population  $\theta$ s were significantly higher for 100 meeting simulations (M = .99, SE  $< .01$ ) than for 25 meeting simulations (M = .86, SE = .04, Mann-Whitney U



(a) Population levels of prior  $P(\text{pun}|\text{theft})$  and  $\theta$  for a representative simulation over generations 1000-4000 (following the onset of  $\theta > 0$  as an evolvable strategy in generation 1000) for the hierarchical model.



(b) Population levels of  $\theta$  (dark blue line) overlaid on a representation of generations with a population  $\gamma > 1.5$  (light blue bars), for the same representative simulation over generations 1000-4000.

Figure 3: Median hierarchical simulation demonstrating cyclic behavior.

test:  $Z = -3.06$ ,  $p < .01$ ). Inspection of the individual simulations revealed that fluctuations of punishment, prosociality and group bias were substantially less frequent in 100 meeting simulations, compared to 25 meeting simulations; however, some fluctuations did occur even with 100 meetings.

### Non-learning control models

All three of the non-learning control models tested yielded similar results: punitive strategies failed to emerge, and innate prior probabilities of punishment tended towards canalized antisociality. The average population thetas on the 4000th generation for the uniform prior model was .01 (SE  $< .01$ ), for the one-level model was .01 (SE  $< .01$ ), and for the hierarchical model was .01 (SE  $< .01$ ). For the one-level model, on the 4000th generation the average population  $\frac{\alpha}{\alpha+\beta}$  was .01 (SE  $< .01$ ) and the average population  $(\alpha + \beta)$  was 1.09 (SE 1.3). For the hierarchical model, on the 4000th generation the average population  $\psi$  was -10.45 (SE 0.73), and the average population  $\gamma$  was 1.09 (SE 1.31).

### Discussion

Our findings suggest that costly punishment of antisocial behavior is adaptively favored when it causes others to act prosocially in future interactions—that is, when agents learn to avoid punishment. This was evident in evolutionary simulations using three cognitive learning models: learning with a fixed uniform prior, a one-level model with evolvable priors, and a hierarchical model with evolvable priors. The evolution of costly punishment was not observed in any of the corresponding non-learning models. It is particularly notable that punishment readily invaded in the uniform prior model, which plausibly approximates the operation of a domain-general learning mechanism (Courville, Daw, & Touretzky, 2006; Tenenbaum et al., 2006). Thus, widely-shared psychological learning mechanisms may be sufficient to render

a selective advantage to the strategy of punishing antisocial behavior.

Additionally, our findings demonstrate that innate behavioral tendencies towards prosocial behavior are adaptively favored in an environment where antisocial behavior is punished. This was evident both in simulations implementing one-level and hierarchical learning models. As population levels of  $\theta$ s changed, innate levels of prosociality changed in response to closely track the true  $P(\text{pun}|\text{theft})$ . These models provide a case-study in the use of Bayesian priors to model innate behavioral dispositions. This approach has attractive characteristics: with a ready set of formal methods for combining innate factors with experienced data, it allows innate factors to vary in their behavioral consequences as well as in their developmental persistence (see also Kirby et al., 2007).

Our findings confirm previous models suggesting that costly punishment and contingent prosociality can coevolve, (e.g. Boyd & Richerson, 1992; Gardner & West, 2004). Indeed, behavioral studies show that people will pay a cost to punish antisocial behavior, which can stabilize prosocial behavior (Fehr & Fischbacher, 2004). However, this same experimental evidence demonstrates that punishment of antisocial behavior by one individual will reliably induce more prosocial behavior when the perpetrator interacts with a different individual. In the most extreme case of such a group-level bias, in which the perpetrator does not differentiate between the punishment rates of individual victims at all, punishment by any individual becomes a public good and should not be selectively favored.

Consequently, we explored the coevolution of punishment and prosociality using a hierarchical Bayesian model of learning in which inferences about an individual's likelihood of punishing depends jointly on the prior history of punishment by that individual and the prior history of punishment by others. Critically, we allowed for mutation and natural selection

in the degree of group bias exhibited by learning agents—that is, their propensity to infer one individual’s punitive response from past experience with others.

The specific set of parameters we tested revealed a striking cyclical dynamic involving innate levels of group bias, punishment, and prosociality. Inspection of the timing of this cyclical dynamic suggests a simple course of events. First, population levels of punishment increase in the population due to the benefits of teaching social partners contingent prosociality. Next, population levels of prosociality increase, reflecting an innate bias towards prosociality that is beneficial in an environment where antisocial behavior is reliably punished. In this environment a strong innate group bias can also emerge, perhaps reflecting the homogeneity of punitive strategies when  $\theta \approx 1$  dominates the population. But when this innate group bias emerges in a population dominated by punishment, it establishes a selective pressure favoring non-punitive strategies (i.e.  $\theta \approx 0$ ). When individuals expect individual punishment because it is frequent at the group level, there is a selective advantage to not punishing, thereby avoiding its costs. Thus, the emergence of a strong group bias in a population dominated by punitive strategies frequently leads to a rapid invasion by the strategy  $\theta \approx 0$ . Following such a shift, population levels of prosociality fall, and then the cycle stands ready to be repeated anew.

The cyclical dynamic revealed in our hierarchical learning model depends on several features of a Bayesian approach to modeling the evolution of social behavior: the ability to capture prosocial behavioral tendencies in terms of innate prior probabilities; the ability to model group bias in the form of a hierarchical learning model; and the ability to flexibly balance innate and learned factors along the dimensions of canalization and prosociality.

However, the choice of a Bayesian framework to represent innate behavioral tendencies carries a definite cost. It has been repeatedly demonstrated in experimental settings that individuals engage in prosocial behavior even when they have no expectation—at least explicitly—of contingent reward or punishment (reviewed in Gintis, Bowles, Boyd, & Fehr, 2005). A standard inference from these findings is that some prosocial behaviors have intrinsic utility (innate or otherwise) beyond the expected value of contingent reward or punishment. Thus, it is unlikely that prosocial behavior can be understood at a mechanistic level just in terms of posterior probabilities of punishment and reward. Nevertheless, the Bayesian framework we adopt here has apparent value in modeling innate behavioral tendencies at a functional level.

Moreover, beyond the specific case of prosocial behavior, it appears that humans are equipped with certain forms of innate knowledge and constraints on knowledge acquisition (Spelke, 2000). The ease with which innate knowledge and constraints can be combined with experience is among the principle benefits of a Bayesian framework. We suggest that a fruitful direction for future research in social cognition will be to introduce these Bayesian models into agent-based simulations

of evolutionary processes, exploring how adaptive pressures shape the boundary between the innate and the learned.

## Acknowledgments

The authors acknowledge the Mind, Brain and Behavior Initiative at Harvard University and AFOSR under a MURI for financial support. We thank Lisa Stewart, Joshua Tenenbaum, David Rand, Chris Barker, Daniel Roy, Whitman Richards and Joshua Greene for their advice and assistance.

## References

- Boyd, R., & Richerson, B. R. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, *13*(3), 171–195.
- Brandt, H., Hauert, C., & Sigmund, K. (2003). Punishment and reputation in spatial public goods games. *Proceedings of the Royal Society: Biological Sciences*, *270*, 1099–1104.
- Courville, A., Daw, N., & Touretzky, D. (2006). Bayesian theories of conditioning in a changing world. *TRENDS in Cognitive Sciences*, *10*(7), 294–300.
- Daw, N. D., & Doya, K. (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology*, *16*(2), 199–204.
- Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *TRENDS in Cognitive Sciences*, *8*(4), 185–190.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140.
- Gardner, A., & West, S. A. (2004). Cooperation and punishment, especially in humans. *American Naturalist*, *6*(6868), 753–764.
- Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2005). Moral sentiments and material interests: Origins, evidence, and consequences. In H. Gintis, S. Bowles, R. Boyd, & E. Fehr (Eds.), *Moral sentiments and material interests: The foundations of cooperation in economic life*. Cambridge, MA: MIT Press.
- Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational cognitive modeling*. Cambridge, UK: Cambridge University Press.
- Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, *104*(12), 5241–5245.
- Sigmund, K., Hauert, C., & Nowak, M. A. (2001). Reward and punishment. *Proceedings of the National Academy of Sciences*, *98*(19), 10757–10762.
- Spelke, L. (2000). Core knowledge. *American Psychologist*, *55*, 1233–1243.
- Tenenbaum, J., Griffiths, T. L., & Kemp, C. (2006). Theory-based bayesian models of inductive learning and reasoning. *TRENDS in Cognitive Sciences*, *10*(7), 309–319.
- Traulsen, A., Pacheco, J., & Nowak, M. (2007). Pairwise comparison and selection temperature in evolutionary game dynamics. *Journal of Theoretical Biology*, *246*(3), 522–529.