DE GRUYTER Open Philosophy 2019; 2: 137–146

Computer Modeling in Philosophy

Rory Smead*, Patrick Forber

Signals and Spite in Fluctuating Populations

https://doi.org/10.1515/opphil-2019-0014
Received April 30, 2019; accepted June 17, 2019

Abstract: Spite (in the biological or evolutionary sense) is behavior that harms others at a cost to the actor. The presence of spite in human and animal populations presents an evolutionary puzzle. Recent work has suggested small populations and pre-play signaling can have a significant effect on the evolution of spite. Here, we use computational methods to explore these factors in fluctuating populations that may go extinct. We find that the presence of spite can make a population significantly more likely to go extinct, but that this does not preclude the possibility of spite reliably evolving. Additionally, we find that the stochastic effects of small fluctuating populations allow for the evolution and predominance of signal-mediated conditional spite across a wide range of conditions. These results suggest that directed harm, even if costly, can play a significant early role in the evolution of social behaviors and this provides a possible origin for punishment.

Keywords: Spite, Game Theory, Signaling, Conditional Behavior, Finite Populations, Punishment

1 Introduction

> *But some men are friends with the whole world in their hearts, and there are others that hate themselves and spread their hatred around like butter on hot bread.*
>
> - Samuel in John Steinbeck's *East of Eden*

Philosophy, as a discipline, has a rich recent history of using formal methods: mathematical logic to analyze language and ontology or probability theory to formalize the relationship between theory and evidence. Computational methods are a recent addition to the philosopher's formal tool kit. These methods harness algorithms to explore systems in ways that outstrip current analytical approaches, such as equilibrium analysis or mathematical proof. In particular, computational approaches have become essential for analyzing evolutionary systems. Taking a dynamical approach to these systems reveals novel behavior[1], and permits the analysis of more complex systems. With evolution, computational methods permit the investigation of more complex arrays of strategies competing in a population, and the introduction of stochastic elements (e.g., drift, varying rates of mutation) into the systems. Any formal model of evolution must make idealizations, and these face ineliminable tradeoffs[2], yet computational methods allow for the exploration of models of increasing complexity. Of course, computational methods carry risks. Increasing the complexity of the model and relying on algorithms for analysis raises the risk that formal artifacts might drive otherwise interesting dynamical behavior. Thus, the construction of these complex models, and the deployment of computational methods, must be handled with sufficient care to guard against this risk.

---

[1] Huttegger and Zollman, "Methodology in biological game theory".
[2] Matthewson and Weisberg, "The structure of tradeoffs in model building".

***Corresponding author: Rory Smead,** Northeastern University, Boston, United States of America;
E-mail: r.smead@northeastern.edu
Patrick Forber, Tufts University, Medford, United States of America

Here we will present an instance of using computational methods in philosophy to study the evolution of social behavior. Earlier work quickly reached the limits of analytical methods, and below we show how to extend that work in new directions by relying on simulation to explore the evolutionary dynamics of complex populations. De-idealizing abstract and idealized models can reveal new insights and explanations into social behavior.

William Hamilton's influential theoretical work identified a paradigm for the evolutionary investigation into different kinds of social behaviors (see Table 1).[3] Hamilton's work is primarily aimed at altruism and cooperation. Yet he also had some interest in spite, and he made a broad claim about the evolutionary prospects of harm. We want to revisit this claim by examining the evolution of spite and conditional harm mediated by pre-play signals. In particular, we explore the role of finite fluctuating populations on these behaviors and show that conditional spite can readily evolve in a wide range of settings. These results suggest that directed harm may play a significant role in the evolution of social behaviors, a role more nuanced than Hamilton envisioned, including as a possible origin for punishment.

Table 1: Classification of behavior based on the fitness effects of the act for both actor and recipient

Actor/Recipient	positive	negative
positive	mutualism	selfishness
negative	altruism	spite

Spite is behavior that harms others at a cost to the actor[4]. It occurs at all levels of biological organization from bacteria[5], to birds[6], to humans[7]. This presents an evolutionary puzzle since it would, by definition, be better not to spite. So why does it evolve? Spite can be represented in an understudied game: the Prisoner's Delight[8] represented in Table 2. The social strategy (not harming) strictly dominates the spiteful strategy and is the unique rational solution to the game.

Table 2: The Prisoner's Delight game where $c > 0$ represents the cost of spite and $b > 0$ represents the harm from spite (i.e., spiteful behavior "removes" this payoff). Payoffs are symmetric and shown only for the row-player. Payoffs have been renormalized to avoid negative payoffs.

	social	**spiteful**
social	$b + c$	c
spiteful	b	0

Evolutionary explanations point to situations in which spite generates *relative* advantage while incurring an *absolute* cost. This can come about via certain population structures[9], greenbeard effects[10], signaling[11],

---

[3] Hamilton, "The genetical evolution of social behavior".
[4] West and Gardner, "Altruism, spite, and greenbeards".
[5] Inglis, et al., "Spite and virulence in the bacterium *Pseudomonas aeruginosa*".
[6] Robinson, "Anti-social and social behaviour of adolescent yellow-rumped caciques".
[7] McAuliffe, et al., "Children reject inequity out of spite".
[8] Skyrms, "Evolution and the social contract".
[9] Hawlena, et al., "The evolution of spite: population structure and bacteriocin-mediated antagonism in two natural populations of *Xenorhabdus* bacteria".
[10] Gardner and West, "Greenbeards". Lehmann, *et al.*, "On the evolution of harming and recognition in finite panmictic and infinite structured populations".
[11] Forber and Smead, "The evolution of spite, recognition, and morality".

or simply small populations[12]. In each case, the key component is *anti-correlation* of strategic behavior—when spite is disproportionally directed at non-spiteful individuals[13]. Let $\hat{r}$ be the degree of anti-correlation in the population. That is, with probability $\hat{r}$ a given type interacts with a different type and with probability $(1 - \hat{r})$ there is random interaction. Let x be the frequency of social types in the population, and $(1 - x)$ be the frequency of spiteful types. Then, the fitness of each type is:

$$F(social,x) = \hat{r}c + (1 - \hat{r})(xb+c), \tag{1}$$

$$F(spiteful,x) = \hat{r}b + (1- \hat{r})xb. \tag{2}$$

Spite will be favored whenever $F(spiteful,x) > F(social,x)$ or:

$$\hat{r} > c/b. \tag{3}$$

This condition is analogous to Hamilton's rule for the evolution of altruism. So, it is possible for spite to evolve. Furthermore, studies have shown that spite may have played an important role in our social evolution[14].

2 Spite in finite fluctuating populations

Finite populations can provide the anti-correlation needed for the evolution of spite. Because individuals do not play themselves, the expected utility from gameplay for an individual i in a randomly mixing population X is:

$$u(i,X) = \Sigma_{i\neq j}\pi(s_i, s_j) / (N-1). \tag{4}$$

Where $\pi(s_i,s_j)$ is the payoff for playing i's strategy against another individual j's strategy and N is the current population size.

In this situation, spite will generate a higher expected utility than social behavior if and only if

$$N<b/c+1. \tag{5}$$

Thus, small populations can favor the evolution of spiteful behavior.

William Hamilton recognized this possibility but gave an argument that it was not of any evolutionary significance:

> A population which is small enough, and sufficiently bunched together, to make possible the distribution of such extensive harm must be in danger already, and the spreading of any strongly spiteful mutation is very likely to cause its extinction. Such trends of selection in small populations, if they occur at all, must act like a final infection that kills failing twigs of the evolutionary tree[15].

More recent work has suggested Hamilton may have been too quick to disregard the effects of finite populations on the possible evolution of spite. Smead and Forber[16] show that it is possible for spite to evolve in finite fluctuating populations. However, that study used a deterministic dynamic that cannot

---

[12] Hamilton, "Selfish and spiteful behavior in an evolutionary model". Smead and Forber, "The evolutionary dynamics of spite in finite populations".

[13] Hamilton, "Selfish and spiteful behavior in an evolutionary model". Skyrms, *Evolution and the social contract*.

[14] Jensen, "Punishment and spite, the dark side of cooperation". Marlowe, et al., "The 'spiteful' origins of human cooperation". Forber and Smead, "The evolution of fairness through spite".

[15] Hamilton, "Selection of selfish and altruistic behavior in some extreme models".

[16] Smead and Forber, "The evolutionary dynamics of spite in finite populations".

account for the potential stochastic effects of evolution in small finite populations. A later study showed spite was possible in small and discrete populations of fixed size with stochastic evolution[17]. What is missing, if we want to directly assess Hamilton's argument, is a study that involves both aspects along with the possibility of extinction. While an analytic approach to modeling would quickly become intractable, computer modeling allows us to explore this very easily.

Suppose that the payoffs from the game directly contribute to an individual's chance of producing offspring. We limit each individual to at most one offspring per generation by restricting the total possible payoff, plus a baseline reproduction chance, to be less than 1. Let R_i be the probability that individual i reproduces during a given generation, then

$$R_i = u(i,X) + w. \tag{6}$$

Where $u(i,X)$ is individual i's expected utility in the current population X and w is a fixed baseline chance of reproducing (which we hold fixed at 0.1 throughout).

To allow for dynamic and fluctuating population sizes, we treat the probability of death independently from the probability of reproducing and as a direct function of the population size and the environmental carrying capacity:

$$D_i = N/\rho. \tag{7}$$

Where D_i is the probability that individual i dies in the current generation, N is the total population and ρ is the carrying capacity of the current environment.

Note that the actual size of the population will depend on the underlying payoffs generated from the game as more efficient strategies generate offspring at higher rates to balance out higher chances of death due to the increased population size. This has two effects. First, larger populations generate less negative assortment, and therefore, less potential advantage for spiteful behavior. Second, smaller populations face an increased risk of going extinct. Moreover, limiting results will be of little use in this context as all populations will eventually go extinct. The infamous quote from Keynes is particularly relevant here: "In the long run we are all dead." This is a trivial analytic result for a population of any finite ρ where payoffs are less than 1. And while this is surely true, that all populations will eventually go extinct, it does not help us understand the evolutionary dynamics in real finite populations. For this reason, we employ computer simulations to explore evolving populations in the short and medium run.

Unsurprisingly, lower carrying capacities increase the risk of extinction. Also, lower costs to spite increase the probability of seeing spite evolve, as expected given previous results. Figure 1 shows the probability of spite coming to predominate the population across a range of different costs. Note that in all simulations presented here and below, unless otherwise stated, we began the populations with a random number from 1 to 10 of each strategy type in the population. We examined the population state after 10,000 generations and considered any population state that had more than 80% of one strategy type as predominated by that strategy.

Extinction is only a real threat with very low carrying capacities ($\rho = 100$ or $\rho = 50$), and in such cases, there does seem to be a tension between the evolution of spite and the threat of extinction. If spite takes over a population, it significantly lowers the effective carrying capacity of that population by reducing overall reproduction rates. This means that extinction only requires a short string of bad luck with more deaths than reproduction.

However, at slightly higher carrying capacities ($\rho = 200$) and with small costs to spite, spiteful behavior can evolve and persist regularly without an overwhelming threat of extinction. Larger carrying capacities ($\rho = 400$) reduce the anti-correlation in finite populations to a marginal amount and the strategic dominance of the social behavior becomes the driving factor in determining the evolutionary outcome.

---

[17] Forber and Smead, "An evolutionary paradox for prosocial behavior".

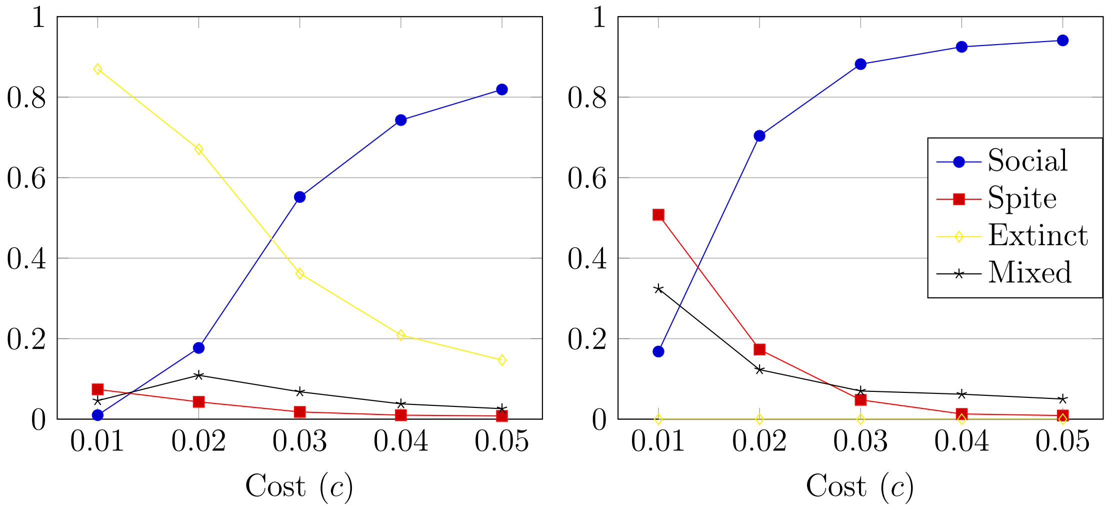

Figure 1: Proportion of simulation runs that are predominated by a particular strategy or go extinct after 10,000 generations for different values of c. For all simulations $b = 0.7$. Left panel shows carrying capacity $\rho = 100$ and right panel shows $\rho = 200$.

Therefore, Hamilton's concern about spite and population extinction is warranted, but only within certain ranges. Spiteful populations are indeed more likely to go extinct. However, this does not entail that spite cannot persist for long periods of time provided that the cost-to-harm ratio is sufficiently small. There is a window of environmental carrying capacities and costs that seem most conducive to producing spite. In this model, the highest proportion of spiteful populations among the cases explored, occurred with $\rho = 200$. If ρ is twice as large social behavior comes to predominate, and if it is half the size spiteful populations tend to go extinct.

3 Signals and conditional spite

Recent studies have begun exploring the connections between spiteful behavior and other forms of costly harmful behavior such as punishment[18]. The origin of punishment is a complex and fascinating topic, rife with conceptual difficulties and evolutionary puzzles that have given rise to a range of different accounts[19]. One intriguing possibility is that some forms of punishment originated as a form of conditionally spiteful behavior[20].

Many of these studies have used strong idealizing assumptions to generate analytic results, and none have explored the possibility of fluctuating population sizes that may be subject to strong stochastic effects in evolution. The model explored above can be expanded to include these features, and computer simulations used to study the effects. We will do this in two ways. First, we will consider a strategy that only employs spite against those that use different strategies. This assumes that there is a reliable way of identifying the strategies of others. To address this, we will consider an elaboration of the model which includes explicit pre-play signaling behavior.

Conditional spite in this game (Table 3) has some interesting properties. It is evolutionarily stable regardless of correlation in the population (both positive and anti-correlation). To see this, note that for positive correlation the strategies performance against itself is overemphasized relative to random interaction and conditional spite does very well against itself. For anti-correlation, the relative payoff differences between strategies of different types are over-emphasized. Again, conditional spite does very

---

[18] Jensen, "Punishment and spite, the dark side of cooperation". Forber and Smead, "The evolution of fairness through spite".
[19] Boyd, et al., "The evolution of altruistic punishment". Nakao and Machery, "The evolution of punishment".
[20] Forber and Smead, "The evolution of spite, recognition, and morality". Smead and Forber, "The coevolution of recognition and social behavior".

well always matching or out performing any opponent in the game. Indeed, if there is sufficient anti-correlation—when condition (3) is satisfied—it becomes the *only* evolutionarily stable state in this game.

Table 3: The Prisoner's Delight game with a conditionally spiteful strategy that only spites different types.

	social	**spiteful**	**cond. spite**
social	$b + c$	c	c
spiteful	b	0	0
cond. spite	b	0	$b + c$

If we examine this game in finite fluctuating populations using simulations, we can see that this conditionally spiteful strategy readily evolves and predominates over a wide range of parameters. We ran 1000 simulations of 10,000 generations for same combination of parameters represented in Figure 1, using random number between 1 and 10 of each type for initial conditions. Conditional spite predominated the resulting populations in a large majority of simulations (the lowest figure was 89.1% with $c = 0.01$ and $\rho = 50$ due to the threat of extinction). This is also true if we consider cases with significantly less favorable b/c ratios. For example, with $c = 0.1$, $b = 0.3$, and $\rho = 400$, 79.7% of populations were predominated by conditional spite after 100,000 generations, a figure that reduces slightly to 74.5% with $\rho = 50$. See Figure 2.

It is also the case that conditional spite evolves if excluded from initial populations and is only introduced via mutation. This takes significantly longer, however. After 10,000 generations with $b = 0.7$, $c = 0.1$, and $\rho = 200$ only 2.4% of populations reached a state predominated by conditional spite. However, after 1 million generations, 61.7% were predominated by conditional spite and the remaining populations were all mixtures of social and conditionally spiteful strategies.

Conditional spite seems to evolve readily and predominate over a wide range of conditions, successfully invading both social and spiteful strategies as well as staving off any extinction threats posed by small populations. However, this assumes that such a conditionally spiteful strategy can perfectly and reliably identify only members of the same type. Thus, the crux is whether or not this conditionally spiteful strategy can emerge spontaneously using signals that are not invariably tied to the behavior. To explore this question, we introduce a signaling element into the game where individuals first send a pre-play signal and then choose a behavior on the basis of the signal received from the opponent. If we focus on just two

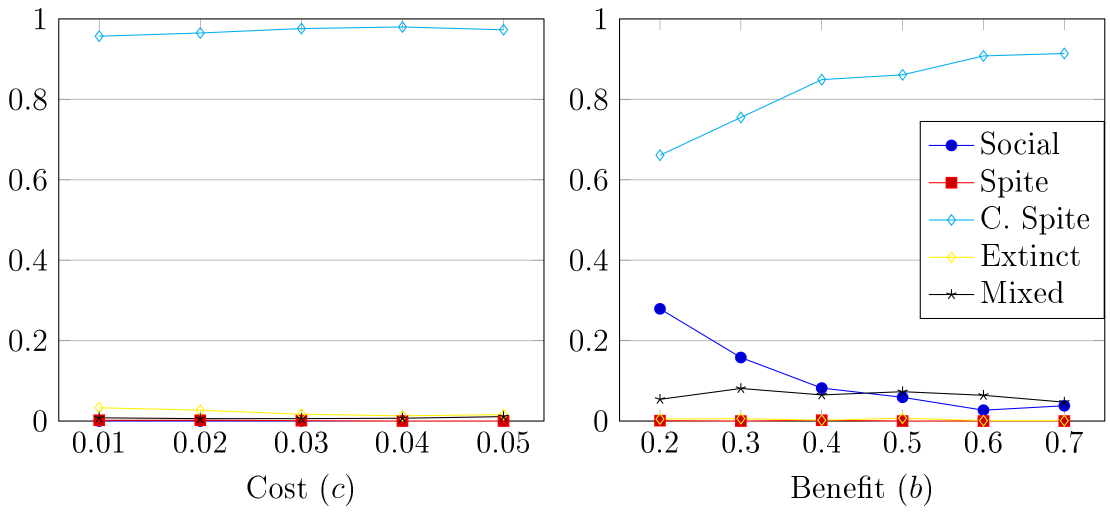

Figure 2: Proportion of simulation runs are predominated by a particular strategy or going extinct as a function of c versus b. For all simulations ρ =100. Left panel shows proportion of simulations for varying costs (with b=0.7); right panel shows proportion of simulations for varying benefits with high cost (c=0.1).

possible signals (a positive signal and a null signal) there are eight possible strategies. We will represent such a strategy with a triple (Signal, Strategy if receiving ‘0’, Strategy if receiving ‘1’). Table 4 shows the full payoff matrix for this game.

Table 4: The Prisoner’s Delight game with pre-play signaling and conditional behavior. Strategies are coded as follows: signal sent (‘0’ or ‘1’), behavior conditional on receiving a ‘0’, behavior conditional on receiving a ‘1’. For example, 1CD sends signal ‘1’, responds to signal ‘0’ with social behavior and to signal ‘1’ with spiteful behavior.

	0CC	0DD	0CD	0DC	1CC	1DD	1CD	1DC
0CC	$b + c$	c	$b + c$	c	$b + c$	c	$b + c$	c
0DD	b	0	b	0	b	0	b	0
0CD	$b + c$	c	$b + c$	c	b	0	b	0
0DC	b	0	b	0	$b + c$	c	$b + c$	c
1CC	$b + c$	c	c	$b + c$	$b + c$	c	c	$b + c$
1DD	b	0	0	b	b	0	0	b
1CD	$b + c$	c	c	$b + c$	b	0	0	b
1DC	b	0	0	b	$b + c$	c	c	$b + c$

Traditional evolutionary analysis shows that there are no evolutionarily stable strategies in this broader game. The reason is that there are neutrally stable sets of strategies that involve a given signal, pure cooperators using that signal, and conditional cooperators using that signal. In infinite randomly mixing populations we should expect some combination of 0CC & 0CD or 1CC & 1DC to evolve with certainty. Those two states are stable with respect to potential invaders of different strategies. Note, however, that with the introduction of constant mutants, the conditionally spiteful strategies should be at a slight disadvantage to their purely social compatriots.

Table 5: The mean population proportions for each strategy in 1000 simulations averaged across the first 100,000 generations. $\rho = 200$. Populations were initialized drawing a random integer from 1-10 as the initial population of each strategy type. Results are rounded to the nearest .001.

condition	0CC	0DD	0CD	0DC	1CC	1DD	1CD	1DC
$c = 0.1$	.279	.001	.199	.002	.294	.001	.002	.223
$c = 0.05$	.191	.002	.321	.002	.180	.002	.002	.300
$c = 0.01$	.047	.180	.256	.022	.046	.172	.022	.255

Simulation results suggest that this slight disadvantage is outweighed by the stochastic effects of selection and the small degree of anti-correlation generated by finite populations. Simulation results under a wide range of parameters show a regular presence of conditionally spiteful strategies (0CD and 1DC). Populations regularly go through phases where these strategies dominate for many generations. See Figure 3.

Spiteful populations still face a threat to extinction with low carrying capacities. In 1,000 simulations with $c = 0.01$ and $\rho = 100$, every population went extinct before 100,000 generations. At $c = 0.01$ and $\rho = 200$ extinction was no longer a threat and populations occasionally showed dominance of purely spiteful behavior, but with transitions to conditional spite occurring regularly. See Figure 4.

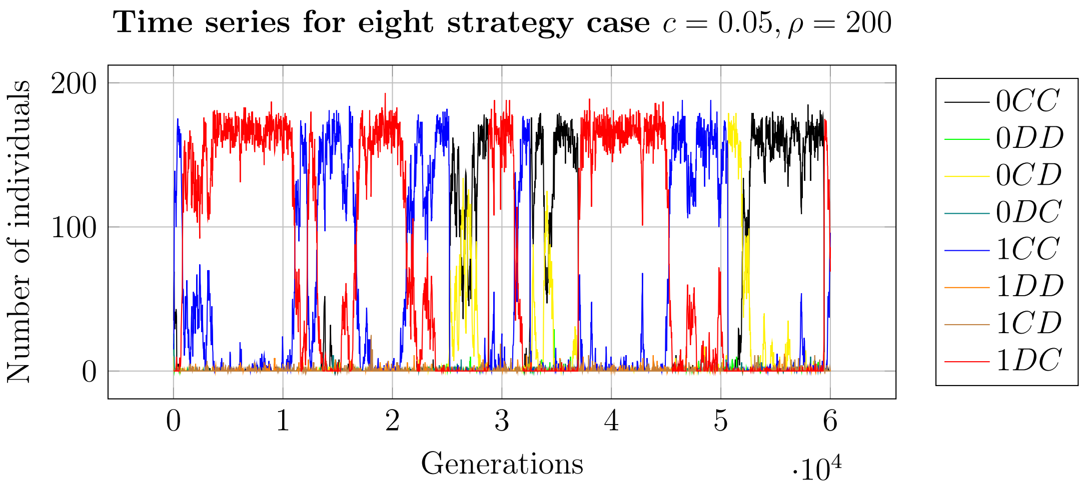

Figure 3: Time series for simulation run showing all eight strategies, $c = 0.05$, $\rho = 200$.

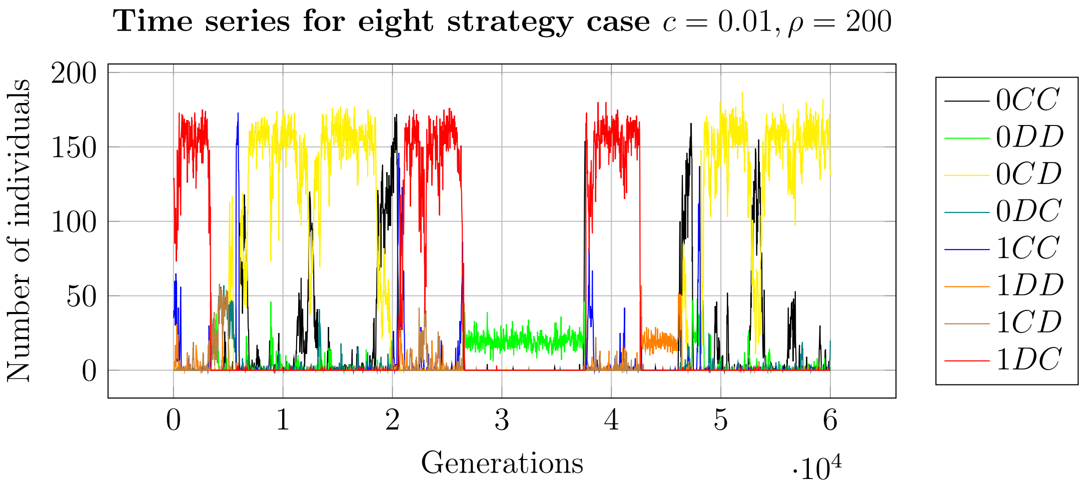

Figure 4: Time series for simulation run showing all eight strategies, c = 0.01, ρ = 200.

With higher relative c values ($c = 0.1$) we no longer see states of abundant purely spiteful behavior. However, populations go through regular transitions between pure social strategies and conditionally spiteful strategies. Occasionally, transitions will occur between signaling states as well.

Conditional spite that uses pre-play signals readily evolves and can predominate in finite fluctuating populations despite having no strategic advantage over non-spiteful strategies. Indeed, they are often at a disadvantage when considering the presence of mutants using other signaling strategies. Despite this, conditional spite evolves due to the stochastic effects of selection and the small degree of anti-correlation generated by finite populations.

4 Discussion

One thread of philosophy of science concerns the structure of possibility space for evolutionary explanation, and how these explanations interact with philosophical questions about fairness, justice, and social structure[21]. With respect to the evolution of social behavior, modeling techniques allow us to investigate

---

[21] Skyrms, *Evolution of the Social Contract*. Skyrms, *The Stag Hunt and the Evolution of Social Structure*. Alexander, *The Structural Evolution of Morality*.

alternative explanations for the social dynamics of cooperation, spite, and punishment. Computational methods are essential to this project because they allow us to investigate more complex arrays of strategies evolving in populations, and they reveal transitions and medium run behavior that would not be revealed by analytical methods such as equilibrium analysis. Using those methods, we demonstrate that signaling and conditional spite can emerge and persist in finite fluctuating populations. This has the potential to refine and reform general explanations for the origins of punishment and the stabilization of social behaviors. The regularity with which conditional spite can arise also suggests that hypotheses about human evolution that presume a robust feedback loop between cooperation and cognitive complexity[22] may be disrupted or complicated by the evolution of directed harm.

These results also have implications for the evolution of punishment. Punishment and conditional spite are both kinds of harmful behavior directed at certain individuals. However, punishment is usually understood to be directed specifically at behaviors and is often related to other social norms[23]. Conditional spite need not be. In some sense, it is more primitive. However, the similarities suggest a potential evolutionary origin for punishment. Our models show that signal-mediated conditional spite can evolve easily through the stochastic effects of finite fluctuating populations. Once present, this behavior provides a basic framework for directing harm toward "outsiders"—those that use a different pre-play signal—which could be interpreted as a type of proto-punishment enforcing group membership. Conditional harm could then be used to enforce different behavior in other games as well[24]. This is not to say that this is *the* explanation of the origin of punishment. Punishment occurs in many forms and many places in the natural world; such a diverse phenomenon presumably has several different explanations. We suggest that punishment via conditional spite is one possible explanation worthy of serious considerations.

Our model also builds on a growing body of research at the intersection of formal evolutionary biology and signaling[25]. For instance, the effects of pre-play signaling have been studied in the Prisoner's Dilemma[26], the Stag Hunt[27] and others. Here we have seen pre-play signaling can have surprising effects in the Prisoner's Delight as well. Lehmann et al.[28] examine the coevolution of neutral markers and conditional social behavior, finding that conditional harm evolves more readily than conditional help—our results lend further credence to this view. Finally, Barrett & Skyrms[29] show how signaling itself can be assembled from degenerate social interactions. This coheres nicely with our approach which shows signals can emerge in unexpected cases.

Let us end on a cautionary note. Computational models have important limitations[30]. Computer simulations allow us to de-idealize abstract models and reveal new explanations[31]. However, they also allow us to construct very complicated models that run the potential risk of introducing artifacts. And, even if we can be ensured that our models are realistic, they may be too complicated to analyze and yield no meaningful understanding of the phenomena they intend to capture. As Borges[32] reminds us, a full-sized map is useless. An overly complicated computer simulation may be equally uninformative.

---

[22] Sterelny, *The Evolved Apprentice*.

[23] Boyd and Richerson, "Punishment allows for the evolution of cooperation (or anything else) in sizable groups".

[24] Forber and Smead, "The evolution of spite, recognition, and morality".

[25] Skyrms, *Signals: Evolution, Learning, and the Flow of Information*.

[26] Robson, "Efficiency in evolutionary games: Darwin, nash and the secret handshake".

[27] Zollman, "Talking to neighbors: The evolution of regional meaning".

[28] Lehmann, et al., "On the evolution of harming and recognition in finite panmictic and infinite structured populations".

[29] Barrett and Skyrms, "Self-assembling games".

[30] Grim, et al., "How simulations fail".

[31] Smead, "The evolution of cooperation in the centipede game with finite populations".

[32] Borges, *Collected fictions*.

References

Alexander, Jason McKenzie. *The Structural Evolution of Morality*. Cambridge University Press, 2007.

Barrett, Jeffrey A. and Brian Skyrms. “Self-assembling games.” *The British Journal for the Philosophy of Science*. 68:2 (2017), 329-353.

Borges, Jorge Luis. *Collected fictions*. Penguin Books New York, 1998.

Boyd, Robert, Herbert Gintis, Samuel Bowles, and Peter J. Richerson. “The evolution of altruistic punishment.” *Proceedings of the National Academy of Sciences*. 100:6 (2003), 3531-3535.

Boyd, Robert, and Peter J. Richerson. “Punishment allows the evolution of cooperation (or anything else) in sizable groups.” *Ethology and sociobiology*. 13:3 (1992), 171-195.

Forber, P and R Smead. “The evolution of fairness through spite.” *Proceedings of the Royal Society B: Biological Sciences*. 281:1780 (2014), 20132439.

Forber, Patrick, and Rory Smead. “An evolutionary paradox for prosocial behavior.” *The Journal of Philosophy*. 111:3 (2014), 151-166.

Forber, Patrick, and Rory Smead. “The Evolution of Spite, Recognition, and Morality.” *Philosophy of Science*. 83:5 (2016), 884-896.

West, Stuart A., and Andy Gardner. “Altruism, spite, and greenbeards.” *Science*. 327:5971 (2010), 1341-1344.

Grim, Patrick, Robert Rosenberger, Adam Rosenfeld, Brian Anderson, and Robb E. Eason. “How simulations fail.” *Synthese*. 190:12 (2013), 2367-2390.

Hamilton, William D. “The genetical evolution of social behaviour. I.” *Journal of Theoretical Biology*. *7*:1 (1964),1-16.

Hamilton, William, D. “Selfish and spiteful behaviour in an evolutionary model.” *Nature*. *228*:5277 (1970), 1218-1220.

Hamilton, W D. “Selection of selfish and altruistic behavior in some extreme models.” In Narrow *Roads of Gene Land: The Collected Papers of W. D. Hamilton Volume 1: Evolution of Social Behaviour*, 198–227. Oxford: Oxford University Press, 1971.

Hawlena, Hadas, Farrah Bashey, and Curtis M. Lively. “The evolution of spite: Population structure and bacteriocin-mediated antagonism in two natural populations of *Xenorhabdus* bacteria.” *Evolution*. 64:11 (2010), 3198-3204.

Huttegger, Simon M. and Kevin J. S. Zollman “Methodology in biological game theory. *The British Journal for the Philosophy of Science*. 64:3 (2003), 637-658.

Inglis, R. Fredrik, Andy Gardner, Pierre Cornelis, and Angus Buckling. “Spite and virulence in the bacterium *Pseudomonas aeruginosa*.” *Proceedings of the National Academy of Sciences*. 106:14 (2009), 5703-5707.

Jensen, Keith. “Punishment and spite, the dark side of cooperation.” *Philosophical Transactions of the Royal Society B: Biological Sciences*. 365:1553 (2010), 2635-2650.

Lehmann, Laurent, Marcus W. Feldman, and François Rousset. “On the evolution of harming and recognition in finite panmictic and infinite structured populations.” *Evolution*. 63:11 (2009). 2896-2913.

Marlowe, Frank W., J. Colette Berbesque, Clark Barrett, Alexander Bolyanatz, Michael Gurven, and David Tracer. “The ‘spiteful’ origins of human cooperation.” *Proceedings of the Royal Society B: Biological Sciences*. 278:1715 (2010). 2159-2164.

Matthewson, John, and Michael Weisberg. “The structure of tradeoffs in model building.” *Synthese*. 170:1 (2009). 169-190.

McAuliffe, Katherine, Peter R. Blake, and Felix Warneken. “Children reject inequity out of spite.” *Biology letters*. 10:12 (2014). 20140743.

Nakao, Hisashi, and Edouard Machery. “The evolution of punishment.” *Biology & Philosophy*. 27:6 (2012), 833-850.

Robinson, Scott K. “Anti-social and social behaviour of adolescent Yellow-rumped Caciques (Icterinae: Cacicus cela).” *Animal Behaviour*. 36:5 (1988), 1482-1495.

Robson, Arthur J. “Efficiency in evolutionary games: Darwin, Nash and the secret handshake.” *Journal of theoretical Biology*. 144:3 (1990), 379-396.

Skyrms, Brian. *Evolution of the social contract*. Cambridge University Press, 2014.

Skyrms, Brian. *The stag hunt and the evolution of social structure*. Cambridge University Press, 2004.

Skyrms, Brian. “Evolution and the social contract.” The Tanner lecture on human values, University of Michigan, 2007.

Skyrms, Brian. *Signals: Evolution, learning, and information*. Oxford University Press, 2010.

Smead, Rory. “The evolution of cooperation in the centipede game with finite populations.” *Philosophy of Science*. 75:2 (2008), 157-177.

Smead, Rory, and Patrick Forber. “The coevolution of recognition and social behavior.” *Scientific reports*. 6 (2016). 25813.

Smead, Rory, and Patrick Forber. “The evolutionary dynamics of spite in finite populations.” *Evolution*. 67:3 (2013), 698-707.

Sterelny, Kim. *The evolved apprentice*. MIT press, 2012.

West, Stuart A., and Andy Gardner. “Altruism, spite, and greenbeards.” *Science*. 327:5971 (2010), 1341-1344.

Zollman, Kevin JS. “Talking to neighbors: The evolution of regional meaning.” *Philosophy of Science*. 72:1 (2005), 69-85.