The Evolution of Spite

Rory Smead and Patrick Forber

DRAFT

November 11, 2011

1 Evolution with correlated interactions

Spite is the shady relative of altruism. Biological altruism is social behavior that incurs a cost (in fitness) to confer a benefit on other individuals. While there are different ways to assess the costs and benefits, and these differences have consequences for the evolutionary dynamics (Kerr et al., 2004), this definition is sufficient to contrast with spite. If altruism involves paying a cost to benefit another, spite involves paying a cost to inflict a greater cost on another. After Hamilton (1964a,b) showed that correlated interactions among kin can produce the evolution of altruistic behavior, he came to realize, along with Price (1970), that spiteful behavior could evolve by a similar process (Hamilton, 1970, 1971). Here we analyze a model of spite (§ 2). Generally, spiteful behavior is unstable, but we show that it can evolve and be maintained under specific conditions. While the model is Hamilton’s, we have a novel extension to cases where population size fluctuates over time. By focusing on a paradigm case of spite, our model helps clarify philosophical puzzles surrounding the concept (§ 3).

1.1 Altruism and correlation

Consider an infinite population of individuals. Let \( x \) represent the frequency of altruists in the population, \( r \) represent the degree of correlated interaction between types in the population (where \( r = 0 \) for random interactions). If \( c \) is the cost of altruism and \( b \) is the benefit to others from altruism, then the fitness functions for altruists \( a \) and egoists \( s \) can be defined as follows:

*Emails: r.smead@neu.edu, patrick.forber@tufts.edu*
The condition for the spread of altruism in a population $x$ is $F(a, x) > F(s, x)$ or:

$$r > \frac{c}{b}.$$ 

The inequality for altruism is: Degree of Correlation > Cost to Altruism / Benefit to Others. This is formally equivalent to Hamilton’s rule – though $r$ need not represent any measure of “relatedness” it is simply some kind of correlation between types. If the above condition is met, altruism will be globally selected in any monotonic selection dynamics.

### 1.2 Spite and anti-correlation

To consider spite, we will imagine a similar situation as above, but now the cost falls on the non-cooperators instead—making the strategy a spiteful one. Suppose now that the population has the following two types of behavior: a beneficial social type $a$ that confers some benefit $b$ at no cost to fitness, and a spiteful type $s$ that pays a cost $c$ to withhold a the benefit $b$ to fitness. With any degree of correlation, spite is not a tenable strategy. Effectively, $c$ is negative in the above condition. This means that in any fixed interaction spite is strictly dominated, reducing the absolute fitness of both parties. However, with some degree of anti-correlation $\bar{r}$ (again, where $\bar{r} = 0$ for random interactions) there can be selection for spite. The fitness functions for $a$ and $s$ can be defined as follows:

$$F(a, x) = (1 - \bar{r})bx,$$

$$F(s, x) = \bar{r}(b - c) + (1 - \bar{r})(bx - c).$$

The condition for the spread of egoism (spite) in a population $x$ is:

$$\bar{r} > \frac{c}{b}.$$ 

The formal parallel to the condition for the spread of altruism is obvious, but it is important to interpret the inequality correctly. In this case the $b$ represents the benefit of normal social interaction, and it is what the spiteful individual denies other individuals. Thus the inequality for spite is: Degree of Anti-Correlation > Cost to Spite / Harm to Others.
2 Spite in finite populations

Small populations can generate anti-correlation, and hence, open the door for spite. Suppose we have a finite but randomly mixing population. Let \( x_t \) represent the number of \( a \)-types at generation \( t \) and \( y_t \) represent the number of \( s \)-types.

Consider the population \( P_t = (x_t, y_t) \) that has the size \( P_t = x_t + y_t \). Suppose the fitness functions are as follows:

\[
F(a, P_t) = \frac{b(x_t - 1)}{x_t + y_t - 1} + k,
\]

\[
F(s, P_t) = \frac{b x_t}{x_t + y_t - 1} + k - c.
\]

\( k \) represents the fitness due to factors outside the interaction; these are assumed to be equivalent among types \( a \) and \( s \). If we let \( F(\cdot) \) denote the expected number of offspring, population size will not necessarily be constant. It is important to note that \( F(\cdot) \) is a measure of absolute (as opposed to relative) fitness. The specifics of the dynamics will be important. We will use an idealized deterministic dynamics that allows for fractions of individuals. This idealization allows for a mathematical simplicity not possible in discrete settings. The dynamics are discrete-time dynamics that allow for changing population sizes:

\[
x_{t+1} = \begin{cases} 
  x_t F(a, P_t), & \text{if } x_t \geq 1 \\
  0, & \text{otherwise}.
\end{cases}
\]

For, the \( s \)-type, we use a analogous function for \( y_{t+1} \). Essentially, the dynamics operate as follows: the number of individuals of a specific type at \( t + 1 \) is equal to the number of individuals at \( t \) multiplied by the fitness. If the fitness is negative, which is possible with a very high \( c \) or when \( x_t < 1 \), that type goes extinct (i.e. the resulting number of individuals is set to zero).

It is worth noting that the dynamics operates on \textit{absolute} fitness values. But if we want to know if spite is increasing in proportion to the rest of the population, relative fitness is what matters. The condition for the relative frequency of type \( s \) to increase is:

\[
\frac{y_{t+1}}{x_{t+1} + y_{t+1}} > \frac{y_t}{x_t + y_t}.
\]

This is equivalent to \( F(s, P_t) > F(a, P_t) \). Given \( P_t \) the population size. The condition above can be reduced to a form that is closely related to the anti-correlation case:
\[ P_t - 1 < \frac{b}{c}. \]

In words the condition says: if the ratio Harm to Others/Cost to Spite is greater than the population size, spite will be favored.\(^1\) Since individuals do not play themselves, as the population size decreases the likelihood that they meet someone different increases; \(\frac{1}{P_t - 1}\) is the degree of anti-correlation generated by the finite population size. The smaller the population, the easier it is for spite to be favored. Hamilton (1970) argued that the small size of populations needed to make spite favored makes them precarious and that this is a reason to doubt the evolutionary significance of spite. We will explore this argument by introducing the possibility of extinction and analyzing the dynamic populations of the model.

The dynamic nature of the population size is essential to understanding the evolution of spite. As the population increases in size the degree of anti-correlation decreases and, eventually, spite will become disfavored. But, as population size decreases, it gets closer to extinction. We will introduce a threshold \(E\) such that if, for some \(t\) \(P_t < E\) the population is considered too small to propagate and goes extinct.

We might now ask: under what parameter values (of \(b, c, k\)) is evolution of spite possible in finite populations? The answer is surprising: almost none. To see why, we first need to consider stability conditions in finite population. If spite is to evolve, it will either dominate the population or persist in a mixed population, hence, there are two cases to consider: monomorphic and polymorphic populations.

### 2.1 Polymorphic populations

A mixed population \(P_t = (x_t, y_t)\) will be considered stable with respect to types if and only if for all \(u > t\):

\[
\frac{x_u}{x_u + y_u} \approx \frac{x_t}{x_t + y_t}.
\]

This means that in \(P_t\) the fitness of each type is equivalent: \(F(s, P_t) \approx F(a, P_t)\). And, since populations size is the only variable factor that can affect \(F(\cdot)\):

\[ P_t - 1 \approx \frac{b}{c}. \]

\(^1\)Hamilton (1971) was the first to derive this condition. Vickery (2003) has a similar finite population model and derives the same condition. Vickery also argues small populations are sufficient for the evolution of spite, but, as we will show below, this is mistaken for populations that are changing in size.
Additionally, $P_t$ to be stable with respect to types, population size must be constant: $P_t \approx P_{t+1}$. To see why, suppose the population size is decreasing, then $F(s, P_{t+1}) > F(a, P_{t+1})$ and since average payoff decreases in this case population size will decrease at a faster rate compounding the effect. Alternatively, if the population size is increasing, then $F(a, P_{t+1}) > F(s, P_{t+1})$ and since this also raises global average payoff, population size will increase at a faster rate compounding the effect. This means that any perturbation in population size will destabilize a population with respect to types.

Furthermore, $P_t \approx P_{t+1}$ and $F(s, P_t) \approx F(a, P_t)$, we know that:

$$\frac{b(x_t - 1)}{P_t - 1} + k \approx \frac{bx_t}{P_t - 1} + k - c \approx 1.$$ 

Which, when using $P_t - 1 = \frac{k}{c}$, can be reduced to:

$$x_t - 1 = \frac{1 - k}{c}.$$ 

Since $\frac{1-k}{c}$ is constant, for $P_t$ to remain stable we need a constant number of $a$-type individuals. If we increase the relative frequency of $a$-types, the population will increase in size the following generation and $a$-types will become favored (and the effect compounded). And if we decrease the relative frequency of $a$-types, the population will decrease in size the following generation and $a$-types will become disfavored (and the effect compounded). Therefore, a perturbation in the relative frequency of $a$-types will also destabilize the $P_t$ with respect to types.

It is reasonable to conclude that evolution will not lead to a stable mixed population since small population perturbations (in type frequency or in population size) can destabilize a population.\footnote{It is important to note that we have not ruled out the possibility of populations that are oscillating around the $P_t - 1 \approx \frac{k}{c}$ point. If $a$-types also oscillate in frequency (low when $P_t - 1 > \frac{k}{c}$ and high when $P_t - 1 < \frac{k}{c}$). The discrete nature of the dynamics allows for this possibility. However, in this case, both population size and frequency of $a$-types must be counter-balanced on each generation for such oscillations. For this reason, we believe that such oscillations are likely to be as fragile as the static situation covered above and hence equally unlikely to occur in evolution. Computer simulations of the model are currently being done which confirm that neither oscillations nor stable mixtures are observed.}

### 2.2 Spite-dominant monomorphic populations

It remains to be seen whether or not a monomorphic population of $s$-types can be stable. To examine this case, we will consider a population $P_t = (0, P_t)$ which consists of all $s$-type individuals where a single $a$-type individual is introduced creating $P'_t = (1, P_t - 1)$.
We will consider \( P_t \) stable if and only if (i) \( F(s, P'_u) \geq F(a, P'_u) \) for all \( u > t \) and (ii) \( P_u \geq 2 \) for all \( u > t \), we can say that spite is stable. The first condition ensures that \( a \)-types are not strongly invasive at any point in the population’s future to the \( s \)-dominant population. The second condition ensures that the \( s \)-dominant population will not go extinct. Let’s suppose \( P'_t < \frac{b}{c} + 1 \) so \( s \)-types are favored in any mixed population. There are three cases to consider:

1. \( k - c < 1 \): The population is decreasing at a rate of \( (k - c) \) per generation: \( P_{t+1} = P_t(k - c) \). The advantage of \( s \)-types continues to increase. Since \( (k - c) < 1 \) and is a constant, \( P_u < E \) for some \( u > t \) and the population goes extinct.

2. \( k - c > 1 \): The population size is increasing at a rate of \( (k - c) \) per generation: \( P_{t+1} = P_t(k - c) \). Since \( (k - c) > 1 \), \( P'_u > \frac{b}{c_1} \) for some \( u \) and type \( a \) becomes strongly invasive: \( F(s, P'_u) < F(a, P'_u) \).

3. \( k - c = 1 \): Here, \( P_t = P_{t+1} \) and there is no change in the population. If \( P_t < \frac{b}{c} + 1 \) then a single invading altruist will be eliminated and the population will increase in size: \( P_{t+1} = P_t\left(\frac{b}{P_t-1} + 1\right) + k \). So long as \( P_{t+1} < \frac{b}{c} + 1 \) this population is still immune to invasion from \( a \)-type individuals. But, if \( P_{t+1} > \frac{b}{c} + 1 \) then the population becomes susceptible to invasion and destabilization. Therefore, the population can be destabilized by a finite sequence of invasion attempts.

Thus, for a spite-dominant population to be stable, it is necessary that \( k - c = 1 \) and that \( P_t < \frac{b}{c} + 1 \). However, this population will not be stable under persistent invasion attempts from \( a \)-type individuals as each attempt increases population size and hence decreases the advantage of \( s \)-types.

### 2.3 Carrying capacities

If the environment has a carrying capacity \( L \), there are new possible avenues for spite to evolve. One way to introduce a carrying capacity into the model above is the following. If any population is such that \( P_t \geq L \) then the population is reduced to \( L \) by eliminating individuals at random. For subsequent generations (in non-shrinking populations) the relative fitness differences drive selection:

\[
x_{t+1} = \frac{F(a, P_t)}{F(a, P_t) + F(s, P_t)} L
\]
It follows that if \( L < \frac{c}{b} + 1 \) and \( k - c > 1 \), spite is guaranteed to evolve since the population is guaranteed to be under the threshold to give spite an evolutionary advantage. Thus, a small carrying capacity relative to the harm to others versus cost of spite ratio can enable the evolution of spite.

2.4 Spite in structured populations

Anti-correlation is what leads to the evolution of spite. Finite populations can generate the necessary anti-correlation, but without some kind of carrying capacity it is unlikely that such populations will be able to allow for the evolution (and persistence) of spite. However, as Skyrms (2007) discusses, anti-correlation can also be introduced by certain population structures.

For instance, certain local interaction settings, where individuals only interact with others nearby, can sustain spiteful behavior (Durrett, 1994; Iwasa et al., 1998). More germane to our model, Skyrms (2007) points out that some group selection models can generate spiteful behavior. He asks us to imagine a population that is divided into small isolated groups (haystacks) where they will remain for a very long time. If the haystacks have a small carrying capacity (below the threshold needed to favor spite) spite will evolve for certain, due to the effects discussed above.

This idea can also be generalized so that the haystacks to not need to persist for very long. Imagine we have a large population \( P - 1 >> \frac{b}{c} \) of individuals that is periodically broken up into a large number of small, equally sized groups \( P_1, ..., P_n \) that are effectively isolated for a few generations. Fitness affecting interactions and reproduction occur within each group for a period of time before being pooled back together, then divided into new groups. Evolution in each group occurs according to our fitness function and population dynamics from above. We will assume that the relative frequency of types in each \( P_i \) is approximately the same as the relative frequency of types in \( P \).\(^3\) Since all groups are the same size and have the same relative frequency of types as \( P \), all \( P_i \) will undergo identical population growth. We also know from above, that if the group size is small enough (\( P_i - 1 < \frac{k}{b} \)) spite will be favored within each group. Thus, as long as the

\( y_{t+1} = \frac{F(s, P_t)}{F(a, P_t) + F(s, P_t)} L \)

\(^3\)These strong assumptions are illustrative but are not strictly necessary for the results discussed here. Similar results are possible as long as (i) group size does not vary too much (ii) there is low variance in the relative frequency of types between groups, and (iii) there are not too many generations spent in the small groups.
groups do not go extinct or grow past the $\frac{b}{c} + 1$ threshold, spite is guaranteed to increase in overall frequency in the population.

Global populations do not need to be small in order to see the spite-inducing effects generated by small populations. It is sufficient that local groups be small. Of course, population structure may introduce anti-correlation that is not due to small population size. On the other hand, any populations that are randomly mixing and not limited in growth are highly unlikely to allow for the evolution of spite. Thus, it seems that some kind of structure (broadly understood) is necessary for the evolution of spite.

3 The concept of spite

The model shows that spite can be stable in populations with sufficiently high $k$ and a carrying capacity. Spite can also exist in declining populations but these populations quickly go extinct. Despite the clear theoretical models of spite, there are claims that spite is just a special kind of altruism or just pure selfishness, and there are a number of difficulties identifying spiteful behavior in the wild. Here we address these conceptual puzzles.

3.1 Spite versus altruism

Spite and altruism are very similar but importantly different. Altruism involves paying a cost to confer some benefit on others. Spite involves paying a cost to inflict a cost on others. Spite and altruism evolve by similar mechanisms. The evolution of altruism requires correlated interaction; the evolution of spite requires anti-correlated interaction.

Lehmann et al. (2006) and West & Gardner (2010) claim that spite simply is a kind of altruism, for inflicting a cost on some members of the population amounts to the same thing as conferring an indirect benefit on the rest of the population. The argument is based on Hamilton’s model for determining the inclusive fitness based on the behavior of a focal individual. Since relative fitness is what matters, a spiteful act can be modeled as inflicting a penalty to receivers of the spiteful behavior, or as conferring an indirect benefit on non-receivers.

This interchangeability works if we assume a constant (or infinite) population size—only relative fitness differences matter. Punishing non-relatives (say) to some degree can yield the same relative fitness differences as helping relatives to a compensatory degree. However, once we let population size fluctuate, spite and altruism come apart.
In our model spiteful individuals pay a cost in terms of absolute fitness to inflict a greater cost in absolute fitness on other individuals, creating a relative advantage. Spiteful behavior decreases group or population productivity. As a result, spite cannot count as altruism on any of the formal definitions of altruism discussed by Kerr et al. (2004). Altruism increases group or population productivity. Furthermore, there is a class of cases where spiteful behavior leads to population decline, whereas altruistic behavior leads to population increase. Consider a small number of migrants joining the population when it near the (unstable) polymorphic equilibrium. If those migrants are a type, the population size increases and number of a types increase, leading to the fixation of a. In contrast, if those migrants are s type, the population will be perturbed in the other direction, leading to the fixation of s, and possibly extinction (see § 2.1). Thus, while spite and altruism can be inter-defined in some cases, they are not equivalent.

Another reason for maintaining the distinction between spite and altruism is that it tracks an ecological difference in animal behaviors. Inflicting a cost is a different causal interaction than conferring a benefit (West & Gardner, 2010, 1344).

3.2 Spite versus selfishness

Many purported examples of spite are often reinterpreted as cases of simple selfishness. Hamilton defined spite as follows: “Let us call an action which harms others without benefiting the self ‘spiteful’, and if a spiteful action involves harm also to self, let us call it ‘strongly spiteful’ (1971, 199).” These costs and benefits are in terms of evolutionary fitness. So, if an organism, during its life history, reaps indirect fitness benefits from apparently spiteful behaviors, then this is simply selfish behavior. West & Gardner (2010) review several purported examples of spite and reinterpret many as cases of selfish behavior by identifying such indirect benefits. The possibility of this reinterpretation illustrates the difficulties of applying the theoretical evolutionary model to real ecological cases.

Identifying the overall costs and benefits, in terms of the evolutionary fitness, for behavior over a complex ecological life history is challenging at best, intractable at worst. Our understanding of fitness and natural selection is mediated by abstract models of a multi-generational process that often make significant idealizations: non-overlapping generations, indefinitely large populations, constant fitnesses, etc. And the fitness properties that are tracked can vary from model to model (see Godfrey-Smith, 2009). Classifying a behavior as spiteful requires committing to a definition of fitness. Suppose we go with
the standard expected number of offspring definition. Then spiteful behavior is behavior that inflicts a penalty on expected number of offspring of others with no benefit to the expected number of offspring for the actor. If the behavior also incurs a cost to the actor in expected number of offspring, as is the case in our model, then it is strongly spiteful behavior.

Clarifying the concept of spite by making explicit the commitments about fitness helps identify real cases of spite. Also, focusing on absolute rather than relative fitness eases the application to the wild. So, for example, if destroying the eggs of conspecifics helps one’s offspring (by decreasing future competition the actor increases the relative advantage in expected number of offspring for her offspring), then this counts as spiteful behavior—the actor’s expected number of offspring does not change. If the actor’s expected number of offspring increases, then the behavior would count as selfish.\(^4\) The key to the evolution of spite is sacrificing absolute fitness to gain an advantage in relative fitness by inflicting a larger cost in absolute fitness on others.

### 3.3 Spite and greenbeards

Gardner & West (2009); West & Gardner (2010) argue that greenbeards can play an important role in the evolution of both altruism and spite. Greenbeard behavior involves a conditional strategy: cooperate with individuals recognized to be the same type; defect (or spite) otherwise. This relies on a recognition mechanism of some sort to generate the correlated interaction necessary for the evolution of altruistic behavior.

First, notice that our model does not require conditional strategies, or kin recognition, to get the evolution of spite. In spite-dominant populations all individuals are engaging in unconditional spiteful behavior. Second, the introduction of conditional strategies changes the strategic interaction, which makes it harder to identify purely spiteful strategies. Populations with conditional spite strategies would exhibit less spiteful behavior overall—individuals would have a conditional strategy of spiting (say) out-group/unrecognized individuals and otherwise interacting normally.

In short, greenbeards provide one possible mechanism for generating correlation or anti-correlation (population viscosity is another), but the introduction of conditional strategies changes the game.

---

\(^4\)It follows that some of the classifications in West & Gardner (2010, 1343, SOM) are ambiguous. For instance, they reinterpret some behaviors as selfish when they benefit offspring. This is either a mistake, or it assumes a longer-term definition of fitness.
References


