## Title

The Iron Law of Selfishness: Response to a comment by Alexander Field



Author
Bergstrom, Ted

Publication Date
2003-05-01

likely to have depends positively upon one's status. This is especially true for males over the span of time over which humans evolved. Stretching a point, status is here interpreted as one's rank in the wealth distribution. Rank matters because one is likely to attract a more fertile mate or mates if one has higher relative income. Frank's view that one's wealth relative to one's peers matters most can also be justified from an evolutionary viewpoint, since competition for mates is usually a "local" matter. Of course, this argument would need to be fleshed out in many ways. However, there is a basis here for an evolutionary justification of risk seeking that fits some empirical facts.

A comparison of the Brenner/Frank view of utility with prospect theory, as developed by Kahnernan and Tversky (1979), is warranted, given the importance of that theory as an alternative to utility theory without relative income as an argument. Prospect theory explains instances of risk seeking, in part, as the result of a utility function that is convex over the range of wealth below the reference point and situations where an individual faces a choice of prospects that all are coded as losses relative to the reference point, which is often interpreted as the prior wealth of the individual making the decision. Thus, the Brenner/Frank utility function uses the wealth of a peer group or of society as the basis for determining relative wealth, while Kahneman and Tversky use the individual's earlier wealth. The two views of the utility function might yield similar predictions in many cases. However, an evolutionary argument for the Kahneman and Tversky (1979) version of reference wealth would seemingly have to be constructed along somewhat different lines than the relatively obvious evolutionary argument for the role of relative income rank in the utility function.

Jon D. Harford
Cleveland State University
Cleveland, Ohio

## References

**Brenner, Reuven.** 1983. *History—The Human Gamble.* Chicago: University of Chicago Press.

**Brenner, Reuven.** 1985. *Betting on Ideas.* Chicago: University of Chicago Press.

**Frank, Robert.** 1985. *Choosing the Right Pond.* Oxford: Oxford University Press.

**Kahneman, D. and A. Tversky.** 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica.* March, 47:2, pp. 263–91.

## Response from Arthur Robson

I am gratified that Neu and Harford share my enthusiasm for developing the biological basis of modern economic behavior. Both of them argue that this approach may lead to preferences that differ from those that are standard in economics. I agree, while also suggesting that an evolutionary perspective should be used to maintain restrictions on the resulting theory. It could then perform "triage" on aspects of economic theory that empirical testing calls into question. That is, an evolutionary perspective could help decide which aspects of this theory should be saved and which should be modified.

Neu argues in particular that an evolutionary approach might favor a "module" theory of mind. Such an approach also seems intrinsically appealing to me, but I expect a challenge will be to retain strong falsifiable consequences of this.

Harford suggests specifically that a biological basis might exist for a nonstandard theory of attitudes to risk where utility depends on rank in the wealth distribution. It is first worth emphasizing that, on pages 95–98 of my article, I showed how the distinction between idiosyncratic and aggregate risk led to attitudes to risk that were interdependent in a related but distinct sense. Furthermore, in Robson (1996), I develop an explicit model that closely follows the pattern he suggests. The model there involves a polygynous mating system, in which the wealth of each male relative to the other males affects his reproductive success. Thus, a concern for status, in this sense, could have been built in. Further, males (but not females) are sometimes risk preferring, despite the Pareto inefficiency of the resulting behavior. Males also tend to favor those fair gambles ("lottery tickets") that have a small probability of a large gain and a large probability of a small loss.

Arthur J. Robson
University of Western Ontario
London, Canada

## Reference

**Robson, Arthur J.** 1996. "The Evolution of Attitudes to Risk: Lottery Tickets and Relative Wealth." *Games and Economic Behavior.* 14:2, pp. 190–207.

## Group Selection and Randomness

In his recent *JEP* article, Theodore Bergstrom (Spring 2002, pp. 67–88) joins a small but distinguished group of economists who have seriously considered the implications of group selection for the conduct of economic inquiry and the assumptions we make about human nature. Until recently, this group has essentially been limited to Gary Becker, Friedrich Hayek, Jack Hirshleifer and Paul Samuelson. Each of these has not only acknowledged the possibility of group level selection (uncontroversial among biologists), but more significantly, has written sympathetically about the possibility that this variant of natural selection has left lasting imprints on human behavioral predispositions.

This note is concerned with one specific claim in Bergstrom's article, however, which I think is wrong or, at best, misleading: the claim that in "haystack" models, group composition must be assortative in order for group selection to attain any traction.

The intuition behind Bergstrom's "theorem" is clear. If altruists interact with defectors, they will have relatively fewer offspring than the defectors at any moment in time. Only if group composition is assortative, requiring that altruists have some way of seeking each other out and differentially associating with each other, can they benefit from their shared altruism and gain in a way that would increase the proportion of altruistic offspring in the population.

In haystack models, populations separate into groups or demes for one or several generations before merging and then reassorting. Bergstrom's logic is premised on the assumption of very large numbers in the individual groups. He argues that if the overall population is large and groups are "formed by random sampling without replacement from this population, then matching will be almost nonassortative" (p. 71). My claim is that the assumption of very large numbers is unrealistic if it is intended to apply to any actual demographic situation under which group selection might apply. Why does this matter? If the numbers are small, random variation alone will almost certainly produce variation in the percentages of altruists within each group. Bergstrom, in fact, acknowledges this: "In haystack models, random group formation produces some groups with more cooperators than others" (p. 71). So long as there is some variation in these frequencies, so that altruists are in some cases grouped together, group selection has the potential to act in a manner that causes the frequency of altruists in the general population to rise. This can happen even if the altruists are declining in each and every group at any moment of time—as long as the groups with relatively more altruists have a greater number of total offspring. And it can happen even though there is no mechanism whereby altruists seek out others similarly inclined and try to join groups differentially composed of them.

My point is related to that used to account for genetic drift, and is based on the statistical properties of small samples. If you flip a true coin, there is a 50:50 chance of getting a head or a tail. It does not follow, however, that if you choose groups of 10, you will always end up with five heads and five tails. The larger the size of the group, of course, the smaller will be the variance of the actual population shares around a mean of a fifty-fifty split. But a variance will remain.

Suppose mutation or genetic recombination has created a small number of altruistically inclined individuals. We are concerned with whether natural selection can allow these genes to persist. Suppose these individuals comprise 10 percent of the total population. Let $n = 100$ and have the population assort periodically into 10 equally sized haystacks. It is quite unlikely that each group will end up with nine defectors and one altruist. Perhaps the ten haystacks would include one that contains three altruistic individuals, one that contains two, five that contain one and three haystacks that contain none. Where small numbers are involved, random variation will produce a variation in trait frequency within groups that produces an outcome that to the untrained eye might in fact look *as if* there had been some tendency for altruists to seek each other out and associate with them. But the process is essentially random.

This point matters because as it stands, Bergstrom's logic appears to require as a precondition for any operation of group selection that individuals be armed with machinery for seeking out and differentially assorting with other cooperators. There is some experimental evidence that we have in fact acquired these capabilities, but to make it a precondition for the evolution of altruism raises unnecessarily the hurdles that group selection must overcome to be considered a potentially serious influence on human nature. It faces enough of these as things stand.

Alexander J. Field
Santa Clara University
Santa Clara, California

## Reference

**Field, Alexander J.** 2001. *Altruistically Inclined? The Behavioral Sciences, Evolutionary Theory, and the Origins of Reciprocity*. Ann Arbor: University of Michigan Press.

## Reply from Theodore Bergstrom

My paper stated and proved a theorem called The Iron Law of Selfishness. The theorem asserts that in a "haystack model" of group formation and dissolution, if (i) groups are formed by nonassortative matching and (ii) at dispersal time the number of descendants of each founder is the payout in a multiplayer prisoners' dilemma game between the cofounders, then the only asymptotically stable outcome is a population consisting of defectors only.

Professor Field maintains that this theorem is "wrong or, at best, misleading," because it "is premised on the assumption of very large numbers in the individual groups." In this, he is mistaken. The theorem does not assume that groups are large, but does rely on groups being populated by random draws from a large overall population. The theorem applies even if colonies are founded by only two individuals.

Professor Field correctly observes that to prove that cooperation will be eliminated, it is not sufficient to establish that the proportion of cooperators in each group declines over the group's lifetime. Since groups are dissolved and randomly reconstituted, he suggests that the high growth rates of those cooperators who are selected into relatively cooperative groups might be sufficient to overcome the disadvantage of lower growth rates within each group. While this speculation has some appeal to plausibility, it does not survive mathematical computation. The Iron Law theorem establishes that if the relevant game between founders is a prisoners' dilemma, then random group formation does not produce enough variation across groups to sustain cooperation.

For many thoughtful people, strong intuitions are not overthrown by a single theorem, especially if the proof is indirect. To build intuition and understanding of a general proposition, it often helps to work through the details of a manageable special case. I will attempt this by proving an Iron Law theorem for two-founder haystacks.

Consider a haystack model with $n$ locations (haystacks), each of which is colonized by exactly two individuals drawn at random from a large population in which the proportion of cooperators is $x$ and the proportion of defectors is $1 - x$. The founders and their descendants live and reproduce (asexually) in their haystacks for a fixed period of time, and then all of the haystacks are dispersed into a single population. The $n$ haystacks are then recolonized by random draws from the population at large, again assigning two individuals to each haystack. At dispersal time, the number of each founder's descendants is given by the payout function in a prisoner's dilemma game. If a haystack is settled by two cooperators, each has $R$ descendants, if it is settled by two defectors, each has $P$ descendants, while if it is settled by one of each type, the defector has $T$ and the cooperator has $S$ descendants. For this to be a prisoners' dilemma, we must have $T > R > P > S$.

With random selection of $2n$ individuals to colonize the $n$ haystacks, the number of cooperators selected is $2xn$ and the number of defectors is $2(1 - x)n$. As Professor Field noted, not all haystacks will be settled by the same proportions of cooperators and defectors. In fact, $x^2 n$ of the $n$ haystacks will be colonized by two cooperators, while $2x(1 - x)n$ will be settled by one cooperator and one defector and $(1 - x)^2 n$ will be settled by two defectors. At dispersal time, haystacks colonized by two cooperators yield $2R$ cooperators, haystacks colonized by one cooperator and one defector yield $S$ cooperators and $T$ defectors and haystacks colonized by two defectors yield $2P$ defectors. Therefore, at dispersal time, the total number of cooperators in the population is $2x^2nR + 2x(1 - x)nS$, and the total number of defectors is $2x(1 - x)nT + 2(1 - x)^2 nP$. The ratio of the total number of cooperators who emerge from the haystacks to the number of cooperators who entered is then

$$\rho_c = \frac{2x^2nR + 2x(1 - x)nS}{2xn} = xR + (1 - x)S,$$

while the ratio of the number of defectors who emerge from the haystacks to the number who entered is

$$\rho_d = \frac{2x(1 - x)nT + 2(1 - x)^2np}{2(1 - x)n}$$

$$= xT + (1 - x)P.$$

When we compare these two growth rates, we find that

$$\rho_d - \rho_c = x(T - R) + (1 - x)(P - S) > 0,$$

where the inequality follows from the fact that in a prisoners' dilemma, $T > R$ and $P > S$. Therefore, after each cycle of group formation, reproduction and dispersal, the proportion of defectors in the overall population rises. If this process is repeated many times, the proportion of defectors in the population will asymptotically approach one.

This proof for the two-founder Iron Law points the way to a proof for founding groups of variable size. The key observation is that defectors will reproduce more rapidly than cooperators if in the dispersed population, the expected number of descendants of a defector exceeds that of a cooperator. Where haystacks are settled by nonassortative random draws from the dispersed population, the probability of being chosen to be a haystack founder is the same for a defector as for a cooperator. Moreover, the probability distribution of the number of cooperators among its cofounders is the same for a founder of either type. But faced with the same group size and number of cooperators among its cofounders, a founding defector will have more descendants than a founding cooperator. Since the expected number of descendants of a representative individual of either type is the sum of the probability of each configuration of cofounders times the corresponding payout, it follows that the expected number of descendants of a defector must exceed that of a cooperator.

The assumption that the game played by founders is a prisoners' dilemma game is more demanding than may first appear. Consider, for example, a haystack model in which several nonoverlapping generations reproduce within each haystack before dispersal. Each individual's number of offspring is determined by a multiperson prisoners' dilemma game played with its own generation in the haystack. In this case, the "game" between founders that determines the number of their descendants at dispersal time is not necessarily a prisoners' dilemma. In fact, Cohen and Eshel (1976) showed that for this model, if the period between dispersals is long enough, a stable population of cooperators can be sustained.

My proof of the Iron Law theorem relies on the assumption that the overall population is large. For small populations, two issues arise. The first is easily dealt with. In a finite population, matching by random selection without replacement does not quite satisfy nonassortativity. This is the case because one's cofounders are drawn from the remaining population, not including oneself. This deference, in a small population, acts against cooperators rather than in their favor, since cooperators draw their cofounders from a less favorable population than do defectors. Thus, even in a small overall population, so long as both types are present, the expected growth rate of the population of cooperators is lower than that of defectors. The second issue is more profound and beyond the scope of this brief discussion. With small numbers, the distribution of groups by their make-up is itself subject to randomness. Realized distributions may be significantly deferent from the expected distribution. In fact, for the haystack process, extinction of either type would be possible simply because the random draws at matching time happened to populate all haystacks with cooperators or all with defectors.

## Reference

**Cohen, D. and I. Eshel.** 1976. "On the Founder Effect and the Evolution of Altruistic Traits." *Theoretical Population Biology*. 10, pp. 276–302.