Altruism, Inclusive Fitness, and "The Logic of Decision"
Author(s): Brian Skyrms
Source: *Philosophy of Science*, Vol. 69, No. S3 (September 2002), pp. S104-S111
Published by: The University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: http://www.jstor.org/stable/10.1086/341838
Accessed: 10/09/2015 17:40

Altruism, Inclusive Fitness, and "The Logic of Decision"

Brian Skyrms†
University of California, Irvine

We show how Richard Jeffrey's *The Logic of Decision* provides the proper formalism for calculating expected fitness for correlated encounters in general. As an illustration, some puzzles about kin selection are resolved.

1. Introduction. It takes more than one symposium to address Dick Jeffrey's contributions to epistemology, Bayesian methodology and decision theory. Two years ago this association convened a symposium on the contributions that he made to decision theory in his book, *The Logic of Decision.* Today, our focus is mainly on Bayesian epistemology, but between Persi Diaconis' talk on probability kinematics and Sandy Zabell's on radical probabilism, I would like to bring *The Logic of Decision* back for a cameo appearance as evolutionary theory.

What I have to say is not really at variance with the most careful treatments of inclusive fitness in the evolutionary literature (for instance Hamilton 1964; Grafen 1982, 1984; Creel 1990; Frank 1998),[1] but it is at variance with many of the expositions of inclusive fitness that one is likely to come across—even in texts and in articles by those who should, and sometimes do, know better.[2] I offer it here as a mini-tutorial from the point of

†Send requests for reprints to the author, Program in Logic and Philosophy of Science, University of California, Irvine, School of Social Sciences, Irvine, CA 92697–5100; e-mail: bskyrms@uci.edu.

1. In fact, a few months after I gave this talk at PSA 2000 I went to a workshop on Groups, Multi-level Selection and Economic Dynamics at the Santa Fe Institute where Theodore Bergstrom presented a paper that gives a much fuller development of the point of view advocated here. I commend his paper, "The Algebra of Assortative Encounters and the Evolution of Cooperation," to the interested reader.

2. Grafen (1982, 1984) identifies offenders.

Philosophy of Science, 69 (September 2002) pp. S104–S111. 0031-8248/2002/69supp-0010$10.00

view of the intersection of decision theory and evolutionary game theory, as a corrective to popular misconceptions and as an illustration of the clarifying power of *The Logic of Decision* in this context.

2. Fitness. Let us suppose that individuals are paired some way or another to play two-person games, that the population is large and that the payoff to that game is reckoned in terms of Darwinian fitness. I will write the property of playing the strategy A as italic *A,* and the property of being paired with a player who plays the strategy A as boldface **A.** The fitness of strategy B played against strategy A is determined by the game and is written as Fit ($B \cap \mathbf{A}_i$). These conventions should make for a compact and easily understood notation provided that the typesetter cooperates.

If individuals are paired at random, then the expected fitness of a strategy is calculated as an unconditional expectation:

$$\text{Fitness (B)} = \text{SUM}_i \text{ pr } (\mathbf{A}_i) \bullet \text{Fit } (B \cap \mathbf{A}_i)$$

Because this resembles the way expected utility is calculated by Savage, we will call this the Savage expectation. As a slogan:

RANDOM PAIRING → SAVAGE

On the other hand, pairing may not be random. There may be assortment of encounters where a strategy is more likely to meet itself, or dissortment of encounters where a strategy is less likely to meet itself, than would be expected on the basis of random pairing. More generally, there may be various kinds of correlation between strategies. In this case, to compute the fitness of a strategy, we need to use conditional pairing proportions instead of unconditional population proportions. As explained in Skyrms (1994), this is like the computation of expected utility in *The Logic of Decision:*

$$\text{Fitness(B)} = \text{SUM}_i \text{ pr}(\mathbf{A}_i|B) \bullet \text{Fit } (B \cap \mathbf{A}_i)$$

The weights of the expectation are here the probabilities of being paired with a strategy conditional on playing the strategy whose expected fitness is being evaluated. As a slogan:

CORRELATION → JEFFREY

3. Altruism. Everyone knows that positive correlation—whatever the cause—is the fundamental basis of the evolution of altruism. For example, "correlation between interactants is necessary if altruism is to receive positive selection" (Hamilton 1975). Hamilton emphasizes that kin selection is only one way of achieving this correlation:

> it makes no difference if altruists settle with altruists because they are related . . . or because they recognize fellow altruists as such, or settle together because of the pleiotropic effect of some gene on habitat preference. (Hamilton 1975, 141; see also the discussion in Skyrms 1996 and in Sober and Wilson 1998)

Perfect self-correlation is perfectly transparent. It means that others do unto you what you do unto them. The effects of less than perfect correlation are quantified by the Jeffrey expectation.

4. Inclusive Fitness. When we use the Jeffrey expectation, we substitute conditional probabilities for unconditional ones. Hamilton suggests an alternative way of looking at the situation, where—at least on the surface—it seems that we retain unconditional probabilities but modify fitnesses by adding a "correction factor" to get "inclusive fitness." What is Hamilton's correction factor? In *must* be given by:

$$\text{HAMILTON} = \text{JEFFREY} - \text{SAVAGE}$$

This is always correct.

5. Inclusive Fitness? There are misleading characterizations of inclusive fitness to be found, even in the writings of those who know better. For an instance, consider Hamilton's (1971) discussion of the Prisoner's Dilemma in "Selection of Selfish and Altruistic Behavior in Some Extreme Models":

> The implications of game-like situations in ecology are not so difficult to see as the implications of the corresponding 'game' in game theory. For example, if Prisoner's Dilemma is played between individuals meeting at random and if the payoffs are fitnesses, we have seen that it 'pays' natural selection to take the selfish course consistently. This is because the type that does so gets greater-than-average fitness when associated with another type, in no matter what ratio. If assortation occurs, however, this outcome is not certain; increasing correlation of partners must eventually reach a point where fitness in the 'homopairs' dominates the mean fitnesses of the types. The concept of inclusive fitness provides a simple test for the resolution of games in this way. **The test consists in adding to the expressed fitnesses a fraction *b* of the fitness of the partner where *b* is the coefficient of relatedness of the partner.** (emphasis mine) The differences in the totals so formed are differences in inclusive fitness. For example, if the partners are sibs
>
> | | | | | |
> |---|---|---|---|---|
> | 2 | 4 | GIVES RISE TO | 3 | 4 1/2 |
> | 1 | 3 | | 3 | 4 1/2 |
>
> showing that with this degree of relationship the incentive to 'let the partner down' has become zero. (1971, 69–70)

One might conclude from reading only this that the emphasized sentence is meant to give a general definition of inclusive fitness. We do find it used as such in the literature, even though a careful reading of Hamilton (1964) should have precluded this.

How does it compare with JEFFREY? Let us suppose that cooperation is a new rare strategy in the population, so that we may approximate the unconditional probability of meeting a cooperator as zero. In Hamilton's example, genetics is diploid and cooperation is dominant. The interactions in question are with sibs. Then the probability of meeting a cooperator given that you are one is equal to1/2, and the probability of meeting a cooperator given that you are a defector is equal to 1. The JEFFREY expectation of cooperation is $1/2*1 + 1/2* 3 = 2$, and the JEFFREY expectation of defection is $1*2 = 2$. JEFFREY agrees that here cooperation and defection has equal fitness.

But the logic of the revised fitness matrix is somewhat obscure. If you interact with your brother, you cooperate and he defects, you add 1/2 his fitness to yours presumably because he has probability 1/2 of carrying the gene for cooperation. But he defected and the gene is dominant, so he does not carry the gene! Shouldn't the Hamilton correction term be added to the expected fitnesses of the strategies, rather than to the individual terms in the payoff matrix?

Furthermore, the Jeffrey expected fitnesses of magnitude 2 are less than any term in Hamilton's revised payoff matrix. If we compute the Savage expectation of Hamilton's revised payoff matrix, assuming that probability of cooperation is zero, we get an expected fitness of magnitude 3 for both cooperation and defection. This suggests some double counting in Hamilton's treatment, but it appears to be innocuous in that it does not affect the qualitative conclusion that expected fitnesses are equal.

But what about a somewhat different payoff matrix, with the interactions being between clones? Using the same procedure:

Defect	2	7	GIVES RISE TO	4	8
Cooperate	1	3		8	6

The modified payoff matrix is indicative of a game where evolution drives a population to a mixed equilibrium. If most cooperate it is better to defect and conversely. This is surely wrong, because if each always meets its own type, cooperation always gets a higher payoff than defection.

6. Inclusive Fitness in 2 × 2 Games. We can resolve the puzzles noted in the foregoing section by explicitly computing the Hamilton correction factor. For purposes of illustration consider the case in which there are only two strategies, A1 and A2. Let the Fitness of A1 played against A1 be a and the Fitness of A1 played against A2 be b. Then:

$$\text{JEFFREY(A1)} = \text{a pr}(\mathbf{A1}|\mathit{A1}) + \text{b pr}(\mathbf{A2}|\mathit{A1})$$

$$= \text{a pr}(\mathbf{A1}) + \text{b pr}(\mathbf{A2}) + \text{a}[\text{pr}(\mathbf{A1}|\mathit{A1}) - \text{pr}(\mathbf{A1})] + \text{b}[\text{pr}(\mathbf{A2}|\mathit{A1}) - \text{pr}(\mathbf{A2})]$$

Let r be $[\text{pr}(\mathbf{A1}|\mathit{A1}) - \text{pr}(\mathbf{A1})]$. Since $\text{pr}(\mathbf{A2}) = 1 - \text{pr}(\mathbf{A1})$ and $\text{pr}(\mathbf{A2}|\mathit{A1}) = 1 - \text{pr}(\mathbf{A1}|\mathit{A1})$:

$$\text{JEFFREY} = \text{SAVAGE} + \text{ar} - \text{br}$$

or

$$\text{HAMILTON} = \text{r(a} - \text{b)}$$

Note the differences (a − b) in the treatment of payoffs and $[\text{pr}(\mathbf{A1}|\mathit{A1}) - \text{pr}(\mathbf{A1})]$ in the relatedness term r. If $\text{pr}(\mathbf{A1})$ and b are both equal to zero we get simple special cases, but in general they will not be zero.

The analysis of the fitness of A2 is similar, with the appropriate payoffs and conditional probabilities. Note that the r used for calculating the fitness of A2 will typically not be the same as the r used in calculating the fitness of A1. Even in the case of perfect assortation $[\text{pr}(\mathbf{A1}|\mathit{A1}) = \text{pr}(\mathbf{A2}|\mathit{A2}) = 1]$ the appropriate r terms for calculating the expected fitness of A1 and of A2 will be equal only if the population is equally split between *A1* types and *A2* types.

The two r terms are not independent, because:

$$\text{pr}(\mathit{A1}) \cdot \text{pr}(\mathbf{A2}|\mathit{A1}) = \text{pr}(\mathit{A1} \cap \mathbf{A2}) = \text{pr}(\mathit{A2} \cap \mathbf{A1}) = \text{pr}(\mathit{A2}) \cdot \text{pr}(\mathbf{A1}|\mathit{A2})$$

The middle inequality is a truth about pairing. For every player who plays A1 and meets a player who plays A2, there is a player—the one he meets—who plays A2 and meets a player who plays A1, and conversely. For example, if both strategies have positive population proportions and one has perfect self-correlation then the other must also.

7. Examples Explained. The foregoing examples can now be treated correctly and the apparent difficulties dissolve. Hamilton gets too high a value for the fitnesses because he multiplies r times absolute payoffs rather than differences in payoffs. Various problems are avoided because of the special nature of the example. If cooperation is a rare mutant, r for cooperation can be 1/2 but r for defection is zero. The fitness of Defection is just its Savage fitness, which is equal to 2. The Savage fitness of cooperation is equal to 1 and the Hamilton correction factor is $(3 - 1)(1/2 - 0) = 1$, so its fitness is also equal to 2.

The second example also ceases to be mysterious if we use differences in computing the Hamilton correction factor. In this example, Savage fitness of defection and Savage fitness of cooperation is 3. If almost all of the population cooperate, r for cooperation is zero and then the Hamilton correction factor for cooperation is also zero. But the coefficient r for defection is equal to 1 and the Hamilton correction factor for defection is −5. Thus the fitness of defection is 2 and that of cooperation is 3 (as we already know from direct computation) and the misleading implications of the modified payoff matrix:

Defect	4	8
Cooperate	8	6

do not arise.

8. Jeffrey's Hamilton's Rule. Suppose that we have an interaction that can be put in this special form:

	Defect	Cooperate
Defect	base	base + benefit
Cooperate	base − cost	base − cost + benefit

We would like to know when the fitness of cooperation exceeds that of defection.

As a consequence of the special nature of the model, the Savage fitness of defection exceeds the Savage expected fitness of cooperation by the cost of cooperation, no matter what the population proportions. If the fitness of cooperation is to exceed that of defection, this deficit must be made up by the respective Hamilton correction factors.

The Hamilton correction factor for cooperation gives an increment in fitness of:

$$[\mathrm{pr}(\mathbf{C}|C) - \mathrm{pr}(\mathbf{C})]\ \mathrm{benefit}$$

and the Hamilton correction factor for defection gives a decrement of fitness of:

$$[\mathrm{pr}(\mathbf{D}|D) - \mathrm{pr}(\mathbf{D})]\ \mathrm{benefit}$$

so the Jeffrey fitness of cooperation exceeds that of defection just in case:

$$[\mathrm{pr}(\mathbf{C}|C) - \mathrm{pr}\ (\mathbf{C}) + \mathrm{pr}(\mathbf{D}|D) - \mathrm{pr}(\mathbf{D})]\ \mathrm{benefit} > \mathrm{cost}$$

The probabilistic quantity in brackets reduces to $[\mathrm{pr}(\mathbf{C}|C) - \mathrm{pr}(\mathbf{C}|D)]$, which we denote by R. (In the case where cooperation is rare, R is approximately equal to the r in the Hamilton correction factor for cooperation.) Thus, cooperation has greater fitness and can spread if:

$$\mathrm{R(benefit)} - \mathrm{cost} > 0.$$

This is Hamilton's rule. The Logic of Decision forces us to make explicit its assumptions.

9. Conclusion. In evolutionary game theory, the real players are the strategies. Strategies persist, while individuals come and go. One can, without doing much harm, think of the average fitness of a strategy as the average fitness of the average individual using that strategy. From this point of view Hamilton (1964) computes this average individual's inclusive fitness by taking her real fitness, subtracting off the increment in expected fitness she got from the cooperation of others, and then adding in the increment in expected fitness she contributed to them (using all the time the proper coefficients r). But here the terms subtracted off and the terms added in are equal! They are both the Hamilton correction factor. This is because only strategies have an identity. If we both are average cooperators, the good I do you must be equal to the good you do me.

We start with Jeffrey expected fitness, subtract off the Hamilton correction factor to get Savage expected fitness, and add on the Hamilton correction factor to get Jeffrey expected fitness back again. *Old-fashioned expected fitness and inclusive fitness are one and the same.* Hamilton (1964) intended just this. Inclusive fitness is not supposed to have a different value, but rather is another way of computing the same value. It has a certain individual altruistic flavor that appealed to Hamilton:

> In essence what I had come to see was the simplification to be effected by attacking the problem from two new points of view. One was simply the 'gene's eye' . . . But, at least as we humans perceive the matter, it is not our genes but *we* . . . that make the decisions, so I had been delighted to find something approaching an individualistic view that I could justify for whole genotypes and which could serve as a guide to social adaptation. This was the idea of inclusive fitness. (Hamilton 1996, 27)

REFERENCES

Bergstrom, T. (forthcoming), "The Algebra of Assortative Encounters and the Evolution of Cooperation", *International Game Theory Review.*

Creel, S. (1990), "How to Measure Inclusive Fitness", *Proceedings of the Royal Society of London* Series B 241: 229–231.

Frank, S. (1998), *Foundations of Social Evolution.* Princeton: Princeton University Press.

Grafen, A. (1982), "How Not to Measure Inclusive Fitness", *Nature* 298: 425–426.

——— (1984), "Natural Selection, Kin Selection, and Group Selection", in J. R. Krebs and N. B. Davies (eds), *Behavioral Ecology: An Evolutionary Approach,* 2nd ed. Blackwell: London, 62–84.

Hamilton, W. D. (1964), "The Genetical Evolution of Social Behavior", *Journal of Theoretical Biology* 7: 1–52.

——— (1971), "Selection of Selfish and Altruistic Behavior in Some Extreme Models", in J. F. Eisenberg and W. S. Dillon (eds), *Man and Beast: Comparative Social Behavior.* Washington, D.C.: Smithsonian Press, 57–91.

——— (1975), "Innate Social Aptitudes of Man: An Approach from Evolutionary Genetics", in R. Fox (ed.), *Biosocial Anthropology.* London:Malaby Press, 133–153.

——— (1996), *Narrow Roads of Gene Land.* Oxford: W. H. Freeman.
Jeffrey, R. (1965), *The Logic of Decision.* New York: McGraw Hill.
——— (1983), *The Logic of Decision,* 2nd ed. revised. Chicago:University of Chicago Press.
Savage, L. J. (1954), *The Foundations of Statistics,* New York: Wiley.
Skyrms, B. (1994), "Darwin Meets 'The Logic of Decision': Correlation in Evolutionary Game Theory", *Philosophy of Science* 61: 503–528.
——— (1996), *Evolution of the Social Contract.* NewYork: Cambridge University Press.
Sober, E. and D. S. Wilson (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior.* Cambridge, Mass.: Harvard University Press.