The Philosophy of Social Evolution

Jonathan Birch

Introduction:

Jumping into the River. . .

In what circumstances does natural selection favour self-sacrifice? As legend has it, the best known answer to this question was first spoken in a London pub, the Orange Tree, that once stood at the corner of Gower Street and Euston Road. One unspecified night in the 1950s, after 'calculating on the back of an envelope for some minutes', the UCL geneticist J. B. S. Haldane remarked to one of his graduate students, John Maynard Smith, that 'he was prepared to lay down his life for eight cousins or two brothers' (Maynard Smith, 1975).

If the story is true, Haldane had latched on to a profound insight about the evolution of cooperation.[1] He had seen that genetically related organisms may, in certain circumstances, have an evolutionary incentive to help one another. He had also seen that the incentive comes in degrees, and that the size of the incentive depends on the closeness of the helper's genetic relationship to the potential beneficiary. Haldane may not have been the first to see these things—R. A. Fisher has a good claim to that accolade—but he may have been the first to glimpse their potential explanatory power.[2] In a 1955 article, he illustrated these ideas with a vivid example:

> Let us suppose that you carry a rare gene which affects your behaviour so that you
>
> jump into a river and save a child, but you have one chance in ten of being drowned,

---

[1]Hamilton vigorously disputed Maynard Smith's account, but eventually came to accept it; see Segerstrale (2013, Chs. 12, 14 and 15). There is no conclusive proof of Haldane's pub quip, although it makes for a lovely story.

[2]In a discussion of why some insects have evolved to be distasteful to predators, Fisher (1930, Ch. 7) observed that, although the trait is unlikely to benefit an eaten individual, it may benefit its siblings; and he writes that 'the selective potency of the avoidance of brothers will of course be only half as great as if the individual itself were protected'. Hamilton, who had read Fisher's book closely as an undergraduate, noted this, along with Haldane's (1955) paper, as an early precursor of his own theory (Hamilton, 1964). In a sense, the whole theory can be seen as a careful unpacking of Fisher's 'of course'. Sometimes the notion of kin selection is projected back even further—on to Darwin—but I regard this as rather tenuous (for reasons set out by Ratnieks and Helanterä 2009 and Ratnieks et al. 2011).

while I do not possess the gene, and stand on the bank and watch the child drown.

If the child is your own child or your brother or sister, there is an even chance that

the child will also have the gene, so five such genes will be saved in children for one

lost in an adult. If you save a grandchild or nephew the advantage is only two and

a half to one. If you only save a first cousin, the effect is very slight. If you try to

save your first cousin once removed the population is more likely to lose this valuable

gene than to gain it. (Haldane, 1955, p. 44)

He went on, however, to suggest that this idea was probably more applicable to insects than to humans:

But on the two occasions when I have pulled possibly drowning people out of the

water (at an infinitesimal risk to myself) I had no time to make such calculations.

Palaeolithic men did not make them. It is clear that genes making for conduct of this

kind would only have a chance of spreading in rather small populations where most

of the children were fairly near relatives of the man who risked his life. It is not easy

to see how, except in small populations, such genes could have been established. Of

course the conditions are even better in a community such as a beehive or an ants'

nest, whose members are all literally brothers and sisters. (Haldane, 1955, p. 44)

Although Haldane never captured these thoughts in a mathematical model, we can see in this verbal description the subtlety of his thinking. He saw a gene for altruism might spread if the benefits fell differentially on other bearers of the gene, but he also realized that a mechanism that relied on individual organisms consciously calculating degrees of kinship was implausible. His proposed solution was 'small populations': if a population is so small that everyone is a close relative of everyone else, the benefits of altruism will tend to fall on genetically similar individuals without the need for any conscious calculation.

The Orange Tree was demolished in the spring of 1963. In the aftermath of the demolition, the Beatles arrived for a photoshoot, and a shot of them leaping above the rubble of Haldane's old haunt would be used as the cover for their 1963 EP, *Twist and Shout*.[3] Elsewhere in London, probably on a bus or in a railway station, a graduate student called Bill Hamilton was making

---

[3]The photoshoot is documented by Schreuders et al. (2008, pp. 55-57). To my knowledge, no one has previously noted the Haldane connection. The site directly adjoins Tolmers Square, where, in the 1970s, a battle raged between developers and squatters over the fate of the old Victorian tenements (Wates, 1976). One of the squatters was George Price (of the 'Price equation'), who lived there in the months prior to his death in 1975 (Harman, 2010).

leaps of a different kind.[4] He was busily preparing a manuscript for submission to the *Journal of Theoretical Biology*, entitled 'The Genetical Evolution of Social Behaviour', which he intended to complement a short note (entitled 'The Evolution of Altruistic Behavior') that had just been accepted by *The American Naturalist*.

Hamilton had independently arrived at a more general form of Haldane's insight, and he had formalized it in a rigorous and detailed mathematical framework. Crucially, he showed that the general sort of process Haldane had described did not require 'rare genes' or 'rather small populations'. Whenever organisms interact differentially with relatives—whether this is due to active kin recognition, or simply due to ecological mechanisms that keep organisms fairly close to their birth site—the conditions are potentially apt for the evolution of altruism. Hamilton also realized that the same framework could extend beyond altruism to explain other kinds of social behaviour: behaviour that was selfish, spiteful, or mutually beneficial. That 1964 paper became one of the most influential in the history of evolutionary theory. The ideas it contained—ideas now known as Hamilton's rule, kin selection and inclusive fitness—changed the way we think about the evolution of social behaviour.

Maynard Smith, by this time a lecturer at UCL, played a pivotal role in getting Hamilton's 1964 paper accepted, and in bringing Hamilton's ideas to a wider audience (Maynard Smith, 1964), but he infuriated Hamilton by attributing the basic idea of kin selection to Haldane (Segerstrale, 2013, Ch. 12). The 1955 article makes it clear that Haldane did conceive of a form of kin selection in an informal and imprecise way, but perhaps not with the sort of quantitative precision he displayed in the alleged pub quip. In any case, there is no doubt that Hamilton deserves the credit—and probably more credit than Maynard Smith gave him—for developing a formal theory of social evolution with that simple idea at its centre.

Hamilton's pioneering work kickstarted a research program now known as social evolution theory—a program in which the concepts of relatedness, kin selection and inclusive fitness continue to play a central role (Frank, 1998; Wenseleers et al., 2010; Bourke, 2011; Marshall, 2015). This is a book about the conceptual foundations and future prospects of that program. I aim to defend the value of Hamilton's basic insights in the face of recent criticism, to clarify a number of thorny issues concerning the structure of social evolution theory as it is today, and to argue

---

[4]Hamilton's institutional affiliation at this time was somewhat ambiguous. His graduate work was funded through a Leverhulme scholarship the Department of Sociology at the LSE, and his 1963 paper carries that address. But he was also part-registered with the Galton Laboratory at UCL, and he gave this as the address for his 1964 paper. He appears to have preferred to work in his rented bedsit in 14 Hadley Gardens, Chiswick, on the top of buses, and in Waterloo station (Segerstrale, 2013).

that the theory, when suitably extended, has the resources to explain phenomena at first glance far removed from the beehive and the ants' nest, including cooperation in microbes, cooperation among the cells of a multicellular organism, and culturally evolved cooperation among the 'Palaeolithic men' (and women) who evidently occupied Haldane's thoughts.

Proximate and ultimate

On 1 February 1961, while Hamilton was struggling with his early models, Ernst Mayr, a professor of Zoology at Harvard, gave a lecture at MIT called 'Cause and Effect in Biology', his contribution to a longer series of lectures on the theme of 'cause and effect' (Mayr, 1961; Lerner, 1965). In that lecture, Mayr drew what would become a highly influential distinction between two types of biological cause.

Underlying Mayr's distinction was the idea that animal behaviour is controlled by a genetic program. This concept of a genetic program leads naturally to a distinction between those causes, such as natural selection, that are responsible for the gradual shaping of genetic programs over evolutionary time; and those causes, such as developmental, physiological, and cognitive mechanisms, that are responsible for decoding and executing the genetic program during the life cycle of a particular organism.[5] Mayr referred to the former as *ultimate causes* and the latter as *proximate causes*. He regarded the former as the proper subject matter of evolutionary biology.

Social evolution theory is concerned with ultimate causes, and so is this book. However, my conception of 'ultimate causes' is somewhat more liberal than Mayr's. Recently, Laland et al. (2011, 2013) have questioned whether the proximate-ultimate distinction is useful at all. They cite processes such as niche construction (Odling-Smee et al., 2003), developmental plasticity (West-Eberhard, 2003), and social learning (Heyes and Galef, 1996; Avital and Jablonka, 2000; Richerson and Boyd, 2005): processes that clearly matter to both evolution and development, but that seem to involve neither the writing nor the execution of a genetic program. I agree with Laland et al. about the importance of these processes, and about the misleading nature of the 'genetic program' concept—and hence of the proximate-ultimate distinction as Mayr conceived it—when these processes are at work. But I see this as a reason to frame the proximate-ultimate distinction in a different way—a way more accommodating of the sorts of processes Laland et

---

[5]See Mayr (1993) for a particularly clear statement of how Mayr understood the proximate-ultimate distinction. The close connection of this distinction to the concept of a genetic program—already very clear in Mayr (1961)—is set out even more transparently here.

al. highlight—rather than a reason to abandon it altogether.

The key, I suggest, is to drop any reference to genetic programs. All we really need, in order to draw a useful proximate-ultimate distinction, is the idea that a behavioural phenotype has an associated *transmissible basis*—a basis which may be partially or even wholly non-genetic—that explains its recurrence across the generations. We can then say that the ultimate causes of a behavioural phenotype are those which explain the origin and maintenance, over evolutionary time, of its transmissible basis in a population of organisms; whereas the proximate causes of a behavioural phenotype are those which explain, in the context of the life cycle of a particular organism, the relationship between the phenotype's transmissible basis and its manifest form.

So, although genes are a very important form of transmissible basis, I think we should allow that in at least some cases—and perhaps especially in the case of humans—non-genetic processes of inheritance, such as cultural and ecological inheritance, also matter, and that the domain of proximate-causal explanation includes questions about how the manifest form of a phenotype relates to these non-genetic transmissible bases. Similarly, although natural selection is a particularly interesting and important cause of the origin and maintenance of the transmissible basis of phenotypes, we should allow that many other processes, including forms of cultural evolution and niche construction, can also be genuine ultimate causes.[6,7]

This pluralistic view about ultimate causes rarely surfaces in Part I of this book, Foundations, which focusses almost exclusively on one particular, much-studied type of ultimate cause: natural selection acting on parentally inherited genetic variation in a constant environment. I focus on this type of cause because it is the focus of Hamilton's work, and the overarching aim of Part I is to clarify the conceptual structure of the theory of social evolution we owe to Hamilton.

In Part II, however, I branch out in various directions, exploring the impact of 'horizontal' (i.e. nonparental) transmission on the genetic (Chapter 6) and cultural (Chapter 8) evolution of cooperation, as well as considering the feedback effects of group size on the evolution of social complexity (Chapter 7). A recurring theme of Part II is the ability of organisms to alter

---

[6]My view here has affinities with the 'extended replicator' view of Sterelny et al. (1996), which also aims to make room for the fact that phenotypic traits can have non-genetic transmissible bases. However, I do not think the transmissible basis of phenotypic traits must take the form of replicators. For example, cultural transmission need not involve replicators (see Chapter 8). For further discussion of non-genetic inheritance, see Mameli (2004); Jablonka and Lamb (2005); Helanterä and Uller (2010).

[7]For further recent discussion of the proximate-ultimate distinction, see e.g. Calcott (2013); Dickins and Barton (2013); Gardner (2013); Haig (2013); Sterelny (2013a); Watt (2013); Otsuka (2015). The debate initiated by Laland et al. (2011) has brought to the surface a remarkable variety of ways in which philosophers and biologists have come to understand Mayr's distinction. I do not intend my own proposal to supplant all others; it is simply how I will construe the distinction for the purposes of this book.

their social neighbourhoods (e.g. through gene mobility, through teaching others, or through promoting the growth of the group) in ways that feed back into the response to selection, an idea Powers et al. (2011) have termed 'social niche construction'. The overall argument of Part II is that Hamilton's ideas are even more powerful, and can explain even more about the natural world, when we relax some of the assumptions about inheritance that characterized his original models.

Foundations and extensions

Here is a brief preview of what is to come. In Part I, Foundations, I aim to construct a coherent picture of the conceptual structure of social evolution theory, a picture that distinguishes the different explanatory roles of three distinct Hamiltonian innovations that are often conflated: Hamilton's rule, kin selection and inclusive fitness. I assemble the picture gradually, focussing in each chapter on a separate key question:

- What are the main categories of social behaviour in the natural world, and how should they be defined? (Chapter 1)

- What is the role of the principle known as 'Hamilton's rule' in explaining social behaviour, and can the rule be defended in the face of recent criticism? (Chapters 2 and 3)

- What is best way to think about the relationship between kin selection and group (or multi-level) selection? (Chapter 4)

- How should we conceptualize an organism's fitness in the context of social evolution? (Chapter 5)

I argue for a set of interlinked answers to these questions. In Chapter 1, I argue that we should re-interpret Hamilton's famous four-part classification of types of social behaviour as a classification based on recent selection history rather than current fitness effects. This is, in effect, to classify behaviours by their function, if one also endorses a recent history account of function. In Chapter 2, I argue that we should think of Hamilton's rule as an *organizing framework* for social evolution research: a framework that helps us compare and interpret the causal explanations of change provided by more detailed models. In Chapter 3, I argue that recent criticisms of the rule, although they do undermine other conceptions of its explanatory

role, do not diminish its value as an organizing framework, and I argue that it still compares favourably to other possible organizing frameworks.

In Chapter 4, I turn to the relationship between kin and group selection, arguing for a proposal—inspired by Hamilton's own brief comments—on which these processes are conceived as varieties of selection on indirect fitness differences, distinguished by their commitments regarding population structure. In Chapter 5, I contrast Hamilton's two alternative conceptions of individual fitness—'neighbour-modulated fitness' and 'inclusive fitness'—and argue that inclusive fitness has distinctive advantages, in so far as it provides a stable criterion for improvement and a standard for optimality in a process of cumulative adaptation.

The overarching message of Part I is that Hamilton's conceptual innovations still provide us with a compelling and explanatorily powerful way of organizing our thinking about the ultimate causes of social behaviour. I hope that, by bringing a degree of clarity to areas in which rival camps of theorists have often talked past each other, this part of the book will help defuse some of the controversies Hamilton's work has provoked in recent years, while at the same time identifying areas in which further productive debate is possible.

As the field of social evolution research has progressed, its explanatory scope has steadily increased, pushing well beyond behavioural ecology's traditional heartland of insects, birds, and mammals to incorporate a hugely disparate range of biological phenomena under the umbrella of the 'social'. As Andrew Bourke (2011, p. 7) notes, the field 'has grown outwards from the study of the beehive and the baboon troop to embrace the entire sweep of biological organization. It claims as its subject matter not just the evolution of social systems narrowly defined, but the evolution of all forms of stable biological grouping, from genomes and eukaryotic unicells to multicellular organisms, animal societies, and interspecific mutualisms'. In the second part of the book, Extensions, I turn to the ways in which recent expansions in the explanatory domain of social evolution theory have generated new conceptual challenges. I ask:

- What are the consequences of horizontal gene transfer for the evolution of cooperation, and for the very concept of relatedness, in the microbial world? (Chapter 6)

- Can social evolution theory shed light on the origins of the complex multicellular life forms, such as plants, animals, and fungi? (Chapter 7)

- Can a concept of cultural relatedness help illuminate the origin and stability of cooperation in humans? (Chapter 8)

In Chapter 6, I argue that horizontal transmission, and the opportunities it creates for altruism-promoting genes to help their potential future bearers, should lead us to revise our concept of relatedness so that it tracks genetic similarity across time, rather than at a single time. In Chapter 7, I defend the idea that we can usefully think of the multicellular organism as a social phenomenon, especially when thinking about the transition from unicellular to multicellular life. I suggest that thinking about this transition in terms of 'the economy of the cell state' can yield distinctive insights into the feedback effects that promote and limit the division of labour among cells. In Chapter 8, I turn to the cultural evolution of human cooperation. I develop a concept of cultural relatedness, and I argue that a cultural version of Hamilton's rule can provide an organizing framework for study of early human social evolution. I propose (and tentatively defend) a speculative 'cultural relatedness hypothesis' regarding the evolution of cooperation in Palaeolithic human populations.

While these 'extensions' may initially seem unrelated, there are important connections. As noted above, the ability of organisms to influence the selection pressures they face is one theme. The concept of relatedness is another. Microbial evolution pushes us to change the way we think about genetic relatedness, while regarding the multicellular organism as a social phenomenon pushes us to take seriously the genetic relatedness that exists within, as well as between, organisms. Human evolution, meanwhile, pushes us to recognize a fundamentally different kind of relatedness made possible by the evolutionarily novel mode of inheritance—culture—our Palaeolithic ancestors mastered. Yet the basic role of relatedness in stabilizing altruism remains essentially the same in all three cases.

The overall message of Part II is optimistic: I argue that the Hamiltonian organizing framework set out in Part I, if suitably revised and expanded, can help us understand far more about the natural world than one might initially suppose. We can make real progress in understanding microbial evolution, evolutionary transitions, and cultural change by viewing them through the lens of social evolution theory, provided the theory is properly understood and adapted where necessary.

This book is a work of philosophy of science.[8] In both parts, the focus is on central theoretical concepts, such as relatedness and inclusive fitness, and on abstract theoretical principles, such

---

[8]To be more precise, the book belongs to a tradition in philosophy of science that engages closely with the theoretical foundations of a scientific discipline and addresses questions specific to that discipline (as opposed to addressing very general questions about, say, the scientific method or the nature of causation). Landmarks in this tradition, from which I have learned a great deal, include Sober (1984); Lloyd (1988); Brandon (1990); Okasha (2006), and Godfrey-Smith (2009b).

as Hamilton's rule and the Price equation. The questions I explore concern how these ideas relate to each other, how they can be used to explain social evolution, and how they can be extended to novel cases. Experimental research is not the main focus of the book. Yet in working on microbial evolution, evolutionary transitions, and cultural evolution, I have found that issues which initially seem to be of purely theoretical or philosophical interest turn out to bring novel, testable hypotheses into view. In Extensions, I put forward several speculative hypotheses I see as interesting and promising. This is not a work of experimental biology, and I have not attempted to verify these hypotheses empirically. But I aim to show that reflecting on the conceptual structure of social evolution theory, by giving us a better understanding of the theory, can also open up new directions for experimental work.

I hope the book can serve as an entry point, for philosophers and biologists, to a range of debates about the conceptual foundations of social evolution theory—some of which have been running for decades, others of which have barely begun. With this in mind, I have assumed no prior knowledge of social evolution theory and tried to avoid inessential mathematical detail. There are a few, fairly self-contained sections containing mathematical arguments (Sections 2.I, 5.II, 6.III, and 8.V): readers are encouraged to work through them, but are also welcome to skip to the key results, which I have tried to indicate clearly.

That said, the book is not a textbook or an introduction. For readers seeking an introduction to the mathematical methods employed in modelling social evolution, or a survey of the empirical literature, there are better books out there.[9] This book is intended as 'one long argument' for the cogency and explanatory power of Hamilton's ideas, not just as a way of understanding natural selection acting on parentally inherited genetic variation in a constant environment, but also as a way of organizing our thinking about the ultimate causes of cooperation among microbes, among the cells of our bodies, and among enculturated human beings.

---

[9]For example, see Marshall (2015) or McElreath and Boyd (2007) for an introduction to mathematical methods, and Bourke (2011) for a synthesis of the empirical literature.