THE NEW FABLE OF THE BEES: MULTILEVEL SELECTION, ADAPTIVE SOCIETIES, AND THE CONCEPT OF SELF INTEREST

David Sloan Wilson

In 1705 Bernard Mandeville published a humorous allegory in verse portraying human society as a bee hive in which every individual is motivated by personal greed but the effect is to make the society hum along as a unit. The following passage conveys the general tone.

> As Sharpers, Parasites, Pimps and Players,
> Pick-pockets, Coiners, Quacks, Sooth-Sayers,
> And all those, that, in Enmity
> With down-right working, cunningly
> Convert to their own Use the Labour
> Of their good-natur'd heedless Neighbour:
> These were called Knaves; but, bar the Name,
> The grave Industrious were the Same.
> All Trades and Places new some Cheat,
> No Calling was without Deceit.

Mandeville's fable of the bees, along with Adam Smith's metaphor of the invisible hand, has long been used to convey the idea that a well-functioning society can be forged out of individual self-interest.[1] This idea has become such a tenet of modern thought that for many it is an unquestioned axiom and for decades it has served as the foundation of formal economic theory.

Evolutionary Psychology and Economic Theory
Advances in Austrian Economics, Volume 7, 201–220

ISSN: 1529-2134/doi:10.1016/S1529-2134(04)07009-7

Since Mandeville wrote his fanciful tale, an authentic scientific theory has arisen that explains how beehives and other adaptive animal societies evolve. This theory has much to say about human societies, but it fundamentally challenges the concept of individual self-interest as we know it. In this essay I will update the fable of the bees based on modern evolutionary theory. The updated version retains Mandeville's emphasis on *self-organization*, in which an adaptive society can operate without any centralized intelligence. However, it rejects the concept of *self-interest* as an adequate description of either the thoughts or actions that enable individuals to self-organize into adaptive societies.

WHAT IS MULTILEVEL SELECTION THEORY?

The Fundamental Problem of Social Life and its Potential Solution

Darwin's theory of natural selection, which explains how individual organisms can become exquisitely adapted to their environments, does not explain the evolution of adaptive societies with equal ease. To understand the nature of the problem, imagine a mutant individual who behaves in a way that increases the survival of everyone in her society, including herself, to an equal degree. Such a "no-cost public good" might not appear very feasible (and will soon be amended), but is useful for illustrative purposes. By increasing the fitness of everyone, the mutant trait will not increase in frequency within the society (other than by drift, which can equally cause a decrease in frequency). This example illustrates the elementary fact that natural selection is based on relative fitness. It's not enough for a mutant trait to increase its own survival and reproduction; it must do so *more* than alternative traits in the population. The relative nature of fitness makes the evolutionary forces within a population insensitive to the welfare of the population as a whole.

If providing a public good requires a private cost, which seems reasonable, then the prognosis for the evolution of adaptive societies becomes even worse. The public benefits count for nothing, no matter how great, while the private cost, no matter how small, causes the mutant type to constitute a smaller fraction of the population. The final outcome is a population devoid of public goods providers, whose members demonically strive to widen their slice of the pie while remaining oblivious to the size of the pie. I have called this the fundamental problem of social life (Wilson, 2002).

Darwin was aware of this problem and proposed a solution. Suppose that a population consists of not one but many social groups. In this case, our mutant no-cost public good provider does not increase her relative fitness within her group, but she does increase the fitness of her group, relative to other groups in the total population. The mutant trait can spread through the total population on the strength of among-group selection, even if it is selectively neutral within groups. If there is a cost of providing the public good, then the fate of the mutant trait will depend upon the relative strength of the opposing forces of within- and among-group selection.

Darwin's potential solution to the fundamental problem of social life is elegant and perhaps even obvious in retrospect. After all, if natural selection is based on relative fitness, it makes sense that group-level adaptation is based on the relative fitness of groups. The modern version of Darwin's idea is called multilevel selection theory (Sober & Wilson, 1998). Very simply, adaptation and natural selection can potentially occur at any level of the biological hierarchy. Adaptation at a given level requires a process of natural selection at that level. It is potentially undermined by lower levels of selection and in turn potentially undermines higher levels of selection. This theoretical framework has the capacity to explain the behavior of individuals who demonically work to undermine their groups (within-group selection), individuals who angelically work on behalf of their groups (the bright side of among-group selection) and avenging angels who work on behalf of their groups to destroy other groups (the dark side of among-group selection). We might not like the dark sides of animal and human nature, but they exist and require a theory to explain them. Even with this briefest of introductions, the discerning reader can probably see how multilevel selection theory has the potential to explain the good, the bad, the beautiful and the ugly.

The Rejection and Revival of Multilevel Selection Theory

In the middle of the 20th century, the likelihood of natural selection above the level of the individual was rejected with such force that the atmosphere of taboo still hangs like a stale odor over current discussions. Nevertheless, this categorical rejection proved to be mistaken, as I and others have documented elsewhere (Michod, 1999; Sober & Wilson, 1998; Wilson, 1998, 1999). Today multilevel selection theory is used to study an extraordinary array of topics, from the origin of life to the nature of religion. I will touch upon a few of the applications to pave the way for our examination of self interest.

One of the most important applications has been termed the major transitions of life (Hammerstein, 2003; Maynard-Smith & Szathmary, 1999; Michod, 1999). Until a few decades ago, evolution was thought to take place entirely by small mutational change. Now a second evolutionary pathway has been identified, in which social groups evolve to be so integrated that they become higher-level organisms in their own right. The single organisms of today, such as you and I, are the social groups of past ages. Each transition, from groups *of* organisms to groups *as* organisms, requires a shift in the balance between levels of selection, restricting the within-group processes that undermine the integrity of the group, making team work the only game in town. For example, imagine a primordial cell in which the genes are independent agents. Some contribute to the economy of the cell, producing public goods that can be used by all. Other genes selfishly replicate, increasing their representation within the cell without contributing to the common good. The evolution of chromosomes neatly solved this problem, by binding the genes into a single structure that replicates as a unit. In the absence of within-cell selection, natural selection becomes concentrated at the between-cell level. This kind of transition has been proposed to explain the origin of life as a society of cooperating molecular reactions, the origin of nucleated cells as a society of bacterial cells, and the origin of multicellular organisms as a society of single cells.

The social insects, including the bees in their hives that inspired Mandeville, are among the newest major transitions of life (Camazine et al., 2001; Seeley, 1995; Seeley & Buhrman, 1999). The industrious honeybees can work to increase the fitness of the colony, relative to other colonies, or to increase their own fitness, relative to other members of the same colony. The reason that social insect colonies can justly be called superorganisms is because among-colony selection dominates within-colony selection. Higher-level selection is strong in part because social insect colonies are initiated by a small number of individuals (minimally one queen fertilized by a single male), reducing genetic variation within groups and increasing it among groups. The theory of kin selection is based on this genetic relatedness and was originally regarded as an alternative to group selection. However, kin selection is now interpreted as a special case of group selection in which high relatedness within groups favors altruism by increasing genetic variation among groups (Hamilton, 1975). Moreover, many social insect colonies are initiated by more than one queen who mate with more than one male. Evolutionary biologists are increasingly turning to social control mechanisms analogous to the chromosome, rather than genetic relatedness, to explain why social insect colonies hum along so well as adaptive units. For example, it would genetically "pay" a honeybee worker to lay her own eggs rather than selflessly raising the offspring of the queen – were it not for the fact that the eggs would be eaten and the deviant worker attacked by other workers. These "policing" behaviors, as they are called, evolved to suppress within-group selection, making between-group selection the only game in town, just like the evolution of the chromosome. Actually, it is an exaggeration to say that between-group selection becomes the only game in town. It is important to stress that none of these policing mechanisms are completely successful. There are rogue genes that avoid the fellowship of chromosomes and rogue bees that manage to dodge the police. Within-group selection happens in these otherwise highly adaptive nonhuman societies, no less than in human social groups.

Multilevel Selection and Human Evolution

Multilevel selection theory and its many applications in the biological world provide a panoramic background for the study of human evolution. To the best of our knowledge, our ancestors lived in nomadic groups of a few dozen to at most a few hundred individuals. These groups merged and divided into a wide variety of groupings for specific activities. Just as we participate in family groups, sports groups, learning groups, fighting groups, decision making groups, and so on, each organized around a specific purpose, our ancestors participated in family groups, gathering groups, hunting groups, raiding groups, and so on. Almost everything was done in a social context; to be alone was to be in grave danger. In every social context, opportunities existed to increase the fitness of oneself relative to others in the same group, or to increase the fitness of one's group relative to other groups. We evolved the behavioral propensities to capitalize on both options. We also evolved the propensity to limit the self-serving behaviors of our social partners, thereby concentrating natural selection at the between-group level. In short, we have evolved the equivalent of chromosomes in cells and policing in honeybees, thereby qualifying at least crudely as the newest major transition of life (Boehm, 1999; Sober & Wilson, 1998).

Modern hunter-gatherer societies and indeed most small human groups exhibit an organization that anthropologist and primatologist Chris Boehm (1993, 1999) has called reverse dominance. Instead of dominant individuals benefiting at the expense of subordinants within their groups, the subordinants are capable of collectively ganging up on would-be dominants. Perhaps this balance of power was achieved through the use of weapons that made everyone lethal or cognitive skills that improved the ability to form alliances. In any case, it resulted in a form of guarded egalitarianism and a quantum jump in the capacity for collective action. Under the constant gaze of their fellows, eternally vigilant against being bossed around, our ancestors were largely constrained to behave in ways that were agreed upon by consensus. The opportunities for widening one's own slice of the pie within the group were not entirely eliminated, but they were severely curtailed. The features that set us apart from all other species, including our capacities for culture, language, and symbolic thought (all communal activities) are increasingly being explained in terms of this shift from primarily within-group to primarily among-group selection.

It might seem that this scenario accords perfectly with the concept of self-interest, since social control makes it in everyone's interest to cooperate. This will soon become the main focus of our attention, but for the moment we need to appreciate that the mechanisms of social control evolve primarily by among-group selection. If I reward you for supplying a public good or punish you for failing to do so, I have provided a second order public good at my own private expense (e.g. Heckathorn, 1990, 1993). Social control systems that cause groups to function adaptively evolve because better controlled groups outcompete less well controlled groups. A good social control system requires time, energy, and risk on the part of individuals and ultimately requires the provision of public goods at private expense, as surely as voluntary acts of altruism (Fehr & Fischbacher, 2002; Fehr & Gachter, 2002; Sober & Wilson, 1998).

Why should these speculations about the distant past concern the modern social theorist? There is a widespread tendency to regard evolution as irrelevant because our behaviors are guided by culture and learning instead of instinct. This view is as mistaken and has the same stale odor of taboo as the rejection of group selection. Learning and the capacity for culture are genetically evolved adaptations that enable us to adapt to our environments faster than genetic evolution alone. They are evolutionary processes in their own right, and like the mammalian immune system, they are elaborately constructed to reach biologically adaptive outcomes. Cultural evolution is as multilevel as genetic evolution (Boyd & Richerson, 1985; Richerson & Boyd, in press). A culturally acquired trait can spread on the strength of within- or between-group selection, just like a genetically acquired trait. Indeed, culture is largely responsible for the human major transition, since acquired traits are much more likely to become uniform within groups and vary among groups than genetic traits. Human history and current events can be regarded as an ongoing process of cultural multilevel selection, whose dynamics are influenced by psychological traits that evolved by genetic multilevel selection in the distant past (Wilson, 2002). Close attention to evolution as a multilevel process is required to understand these complex but ultimately comprehensible subjects.

At the formal theoretical level, multilevel selection theory has two virtues. First, it provides a complete accounting system for evolutionary change. To build a multilevel selection model, one must specify a global population that consists of a number of local groups. The dynamics within groups, the manner in which groups arise and disappear, and other parameters that define the structure of the global population must all be specified. Only then can global evolutionary change be monitored as a combination of within- and among-group processes. As we shall see, all models of social evolution must assume a multi-group population structure but in many cases the assumptions are not made explicit and the models do not provide a complete accounting system for evolutionary change.

The second virtue is that the distinction between levels of selection captures the essence of what is required for groups to evolve into adaptive units. If we merely want to know if a given trait evolves, it is sufficient to calculate its change in frequency in the global population, without caring whether its advantage resides within or among groups. If we want to ask the more focused question of when groups evolve into adaptive units, then it becomes critical to partition global evolutionary change into within- and among-group components. Multilevel selection theory carves the evolutionary process at its natural joints for the particular question that Mandeville was addressing with his fable of the bees.

It is important to stress that this multilevel framework needs to be employed for all traits associated with human sociality – not just conventional altruism but also social control, cultural processes involving imitation and learning, and traits conventionally regarded as self-interested in a benign sense. In all cases the question to ask is “Does trait *x* replace alternative traits by virtue of a relative fitness advantage *within* groups or *between* groups?” When this nested series of fitness comparisons is carefully employed, the set of group-level adaptations includes but extends far beyond traits that are conventionally regarded as altruistic.

WHAT IS INDIVIDUAL SELF-INTEREST?

Multilevel selection theory explains adaptive groupings as a product of *group-level selection*. Mandeville and the tradition that he represents explain adaptive groupings as a product of *individual self-interest.* On the surface, it would be difficult to imagine two positions more opposed to each other.

To proceed further we must understand what is meant by the concept of individual self-interest. One possibility is that it will provide an *alternative theory* that explains adaptive groupings better than multilevel selection theory. Another possibility is that it provides an *alternative perspective*, that explains adaptive groupings only when multilevel selection theory arrives at the same conclusion. A third possibility is that it simply fails to provide a coherent account of adaptive groupings.

Unfortunately, there is no single “it” when it comes to the concept of individual self-interest, but rather a number of concepts that are rarely distinguished and often incompatible with each other. At least three layers of diversity can be identified.

(1) The evolutionary concept of self-interest is based on survival and reproduction, but it includes a number of different formulations, as I will show.

(2) The economic concept of self-interest is based on maximizing a utility that is often unspecified and therefore broader than the evolutionary concept of self-interest, with many different specific formulations.

(3) Self-interest can be defined in terms of behavioral actions or psychological motives. In evolutionary biology this distinction is nicely captured by the more general distinction between ultimate and proximate causation. All adaptations must be explained in terms of their effects on survival and reproduction (behavioral actions) but also in terms of the proximate mechanisms that cause organisms to act as they do (psychological motives). These explanations are complementary and one can never substitute for the other. Economic models often ignore this complementarity by including psychological and material costs and benefits in a single utility function.

In this essay I will concentrate largely on evolutionary concepts of self-interest (1) and the proximate/ultimate distinction (3). However, I doubt that the expanded range of utilities considered by economic models will alter my basic conclusions. Let us begin by considering the diversity of concepts of self-interest that exist within evolutionary biology.

Self-Interest as Maximizing Relative Fitness Within Groups

Within the framework of multilevel selection theory, it is easiest to associate self-interest and group-interest with within- and among-group selection respectively. Thus, purely self-interested individuals are a product of pure within-group selection and are driven exclusively to maximize their relative fitness within groups. Purely group-interested individuals are a product of pure among-group selection and are driven exclusively to maximize the fitness of their group, relative to other groups in the total population. It is obvious that by these definitions, the claim that adaptive groups can be forged out of self-interest is as wrong as it can possibly be. The essential insight of multilevel selection theory is that natural selection within groups is insensitive to the welfare of the group, as we have seen. If a disposition to behave that evolves entirely by within-group selection benefits the whole group, it does so only as a coincidental byproduct. To claim that relative fitness maximization within groups miraculously results in adaptive groups is to ignore the reasoning that led to the rejection of group selection in the 1960s and all the subsequent developments that led to its revival.

Self-Interest as Maximizing Absolute Individual Fitness

Despite the importance of relative fitness comparisons in multilevel selection theory, the everyday concept of self-interest is often framed in absolute terms. Purely self-interested individuals are driven to maximize their welfare without regard to others, not in comparison to others. If I help you in the course of helping myself, that is no better or worse than if I hurt you, since my only goal is my own welfare, construed atomistically.

At the formal theoretical level, most economic models and many evolutionary models of self-interest are similarly based on the maximization of an absolute quantity. In economic models the quantity is a utility function whose details are often left unspecified. In evolutionary models the quantity is fitness, defined in terms of survival and reproduction. In both cases, the models axiomatically assume that the absolute value of the quantity will be maximized. Economists take it as a given that individuals are utility maximizers. Similarly, when an evolutionary biologist predicts that an animal will cooperate rather than defect because its fitness as a cooperator is higher than its fitness as a defector, the evolutionary biologist is assuming that the animal has evolved to maximize its absolute fitness (Wilson, 2004).

What justifies this assumption? All evolutionary biologists acknowledge that natural selection is based on relative fitness, so the absolute fitness criterion must be justified in terms of relative fitness. This is the case for nonsocial traits that influence only the fitness of the actor. In a population with two types, A and S, if A's absolute fitness is higher than S then its relative fitness is higher as well. However, the situation becomes more complicated for traits that influence the fitness of others in addition to the actor, as we have already seen. If the population consists of a single group and A-types are public good providers that increase everyone's fitness (including their own) by x units at a private cost of y units, then their relative fitness is lower and they will decrease in frequency despite having increased their absolute fitness whenever $x - y > 0$. In this case the absolute fitness criterion cannot be justified in terms of relative fitness and fails to correctly predict what evolves.

Continuing this example, suppose that our A and S types live in an infinite population randomly subdivided into groups of size N within which the public goods are shared. A-types are less fit than S-types within every group, but groups with more A-types contribute more to the total gene pool than groups with fewer A-types. It turns out that for these particular assumptions, the trait evolves by among-group selection, despite being selectively disadvantageous within groups, whenever $x - y > 0$; that is, whenever A-types increase their own absolute fitness. Another way to state the same result is by noting that in randomly formed groups, an A-type's effect on itself $(x - y)$ causes evolutionary change but its effect on other members of its group does not because they are just a random draw from an infinite population (Grafen, 1984; Nunney, 1985).

To summarize, for traits that influence the fitness of others in addition to the actor, the absolute fitness criterion can be justified in terms of relative fitness, but only given certain assumptions about the population structure. If there is only one group or if variation in the frequency of A among groups is either above or below random, then self-sacrificing or spiteful traits that violate the absolute fitness criterion can evolve. In addition, very plausible social interactions can cause the absolute fitness criterion to fail even in randomly formed groups. Since this point is not widely appreciated, I will illustrate it with two elaborations of our example involving public good providing A-types and freeloading S-types (see Wilson, 1998, 2004; Wilson & Kniffin, 1999 for extended treatments).

In the first elaboration, instead of just two types consider a range of strategies of the form "act as an A-type if there are fewer than x other individuals acting as an A-type in your group," in which x varies from 0 to N, where N is the size of the group. This range of strategies includes the original unconditional S-type ($x = 0$) and unconditional A-type ($x = N$), but the inclusion of the intermediate types alters the outcome of the model by causing behavioral variation among groups to be less than random, even in randomly formed groups. This happens because the intermediate strategies tend to compensate for each other, with those less sensitive to providing the public good "turning off" in the presence of those more sensitive. A group will be entirely altruistic only when it consists entirely of the $x = N$ type, because every other type reverts to selfishness in the presence of a sufficient number of altruists. Similarly, a group will be entirely selfish only when it consists entirely of the $x = 0$ type, because every other type reverts to altruism in the presence of a sufficient number of selfish individuals. The below-random *behavioral* variation among groups caused by these social interactions has an effect on the evolution of public good provision similar to below-random *genetic* variation in the original model.

In the second elaboration, suppose that A and S are socially transmitted behaviors rather than genetically innate. However, the process of social transmission is governed by innate rules that evolve by genetic evolution. Groups are formed at random by individuals bearing their innate transition rules and an acquired behavior (A or S) from their previous interactions. The behavioral composition of the groups then changes according to the transmission rules. Under plausible conditions, transmission rules can evolve that create above-random behavioral variation, even in randomly formed groups. All of us are familiar with the internal and external pressures to conform, promoting uniformity within groups and differences among groups. The above-random *behavioral* variation among groups caused by these social interactions has an effect comparable to above random *genetic* variation in the original model.

How do these elaborations escape the argument that effects on others in randomly formed groups have no effect on global evolutionary change? This argument assumes that no one else in the group will respond to the actions of the focal individual. If others either withdraw or contribute their own social benefits in reaction to the focal individual's action, then the calculation of costs and benefits that ignores the response becomes invalid. For example, imagine an individual who decides to provide a public good because his share of it exceeds his private cost ($x - y > 0$). That individual will be disappointed if another member of his group stops contributing to the public good based on his decision, leaving him stuck with the cost ($-y$). Similarly, an individual who decides not to provide a public good because his share does not exceed his private cost ($x - y < 0$) has made the wrong choice if others would have provided public goods on the basis of his decision.

In short, the absolute fitness criterion can be justified in terms of relative fitness only for a narrow range of parameter values. Departures from randomness at the behavioral level can invalidate the absolute fitness criterion, as surely as departures from randomness at the genetic level. This fact is not widely appreciated because the core evolutionary models of social behavior (such as kin selection) assume that behaviors are coded directly by genes, thereby ignoring the many plausible mechanisms that can generate nonrandom behavioral variation in randomly formed groups.

Let us now take stock of the absolute fitness criterion. Models that employ the criterion must specify the same basic parameters as multilevel selection models – a global population that consists of a number of local groups, the dynamics within groups, the manner in which the groups arise and disappear, and so on. Nothing new has been added – these are properties of the social environment, not any particular theoretical framework. The absolute fitness criterion is a claim about what evolves in a particular social environment. As we have seen, the claim is not general but holds only for a narrow range of parameter values. Earlier I stated that multilevel selection theory has the virtue of being a complete accounting system for evolutionary change. The absolute fitness criterion does not possess this virtue.

Even when the absolute fitness criterion can be justified in terms of relative fitness, it only accounts for what evolves in the total population. More is required to address the specific question of when groups evolve into adaptive units. According to multilevel selection theory, this happens precisely when among-group selection outweighs within-group selection. Behaviors that satisfy the absolute fitness criterion can evolve by either within- or among-group selection – in the later case, when the individual's share in the public good outweighs the private cost. When we use the absolute fitness criterion to ask the central question "can adaptive groups be formed out of self-interested individuals?", the answer is "yes," but only under the conditions already specified by multilevel selection theory; i.e. when among-group selection outweighs within-group selection. The concept of self-interest per se adds nothing.

This point can be illustrated with an example from the social insects – the real life equivalent of Mandeville's fable. In the desert fungus-growing ant

Acromyrmex versicolor, colonies are initiated by groups of queens rather than by single individuals. The queens must do the work of the colony until the first generation of workers is born. One queen becomes a specialized above-ground forager, gathering vegetation for the fungus garden, while the others remain below ground to excavate the tunnels and cultivate the garden. Foraging above ground is more dangerous than staying below ground so the forager substantially reduces her fitness compared to the other queens in her colony (Rissing et al., 1989; Wilson, 1990). How can we explain this apparent example of altruistic behavior? The first impulse of biologists was to assume that the queens are genetically related, but DNA evidence revealed that the groups are randomly formed. The next impulse was to assume that the forager was forced to assume her role on the basis of aggressive interactions, but behavioral observations revealed no evidence for fighting and the forager is not smaller than the other queens. The correct answer appears to involve a very strong form of among-group selection. Many colonies are initiated at the same time and the first to raise a generation of workers succeeds at becoming established in competition with the other colonies. If a specialized above-ground forager provides a sufficient competitive edge in among-group competition, the behavior can evolve despite its selective disadvantage within groups – even when the groups are randomly formed. Knowing all of this, we can imagine the specialized forager reasoning about her options as if she were a person who cares only about her absolute fitness: “If I don’t forage, I will be safer but my colony will lose in competition with other colonies. If I do forage, I might die but if I live my colony will prevail against other colonies and I will share their success. Therefore I will forage based only upon my self-interest. The fact that I have reduced my fitness compared to the other queens in my colony is irrelevant, because they are just a random draw from the total population.” This reasoning correctly predicts that above-ground foraging will evolve, but only because strong within-group selection is counterbalanced by even stronger among-group selection. Self-interest does not provide an explanation of the behavior that stands as an alternative to group selection. Instead, group selection is merely folded into the definition of self-interest. To conclude, the absolute fitness criterion adds nothing to the essential insight of multilevel selection theory, which is that adaptive groupings evolve by among-group selection.

Inclusive Fitness and Other Extended Definitions of Self-Interest

When groups are non-randomly formed, effects on others – now a biased sample of the total population – produce evolutionary change in addition to effects on oneself. A common way to model these effects is with a self-term and an other-term appropriately weighted by a coefficient. In kin selection theory the coefficient is associated with the probability of sharing genes identical by descent, but more generally it can be interpreted as a covariance between the individual's genotype and the genetic composition of its group (positive for above-random and negative for below-random variation).

It might seem that these models straightforwardly show how self-interest is not sufficient to explain the outcome of evolution whenever the covariance term is nonzero. However, a shift in perspective enables both the self and other terms to be interpreted as components of an expanded definition of self-interest. In kin selection theory, individuals are said to maximize their own inclusive fitness, which is often explained as an interest in their own genes, regardless of the bodies in which they exist. In selfish gene theory, the self-interested agent is shifted from individuals maximizing their inclusive fitness to the genes themselves.

Figure 1 shows how any trait that evolves in a multi-group population can be interpreted as a form of self-interest, no matter how self-sacrificial (from Sober & Wilson, 1998). The total population consists of two types, A and S, subdivided into two groups of size $n = 100$ (smaller circles). Both types have a baseline fitness

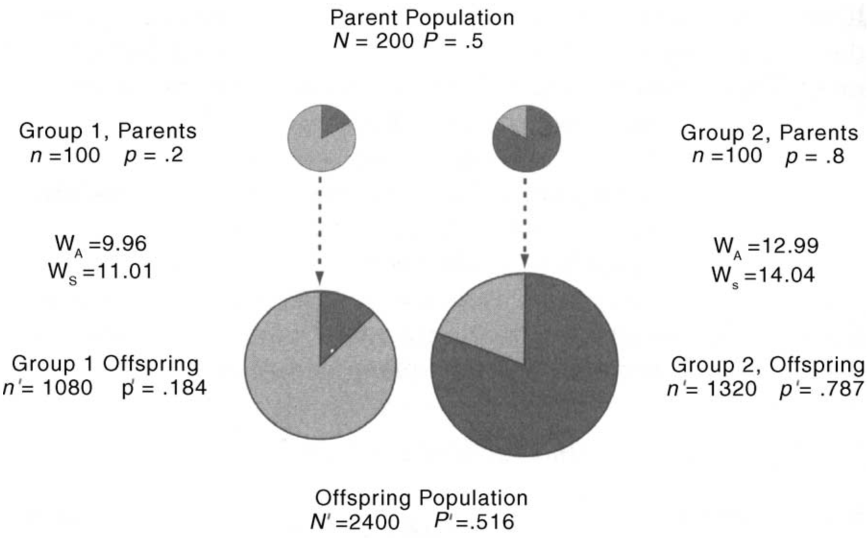

Fig. 1. Evolution in a Two-Group Population with an Altruistic A Type (Dark Shading) and a Non-Altruistic S Type (Light Shading). *Note:* Altruism is selectively disadvantageous within each group but favored sufficiently at the group level that the global frequency of a increases from 0.5 to 0.516.

of 10 and A-types increase the fitness of a single recipient in their group by five units at a cost to their own fitness of one unit. They decrease their own absolute fitness, which means that this behavior would not evolve in randomly formed groups. However, the groups in Fig. 1 are highly segregated, with A at a frequency of $p = 0.2$ in group 1 and $p = 0.8$ in group 2. A-types have a lower fitness than S-types and decline in frequency within each group (within-group selection). However, the group with more A-types grows larger and contributes more to the total gene pool than the group with fewer A-types (between-group selection). The balance between levels of selection is such that A-types increase in frequency in the global population, despite decreasing in frequency within each group. Extreme variation among groups is required to produce this outcome. For real-world examples we must specify how the variation is created and maintained, especially at high and low frequencies of the altruist in the global population. Interesting forms of frequency-dependence can occur resulting in stable polymorphisms of the two types, as discussed by Sober and Wilson (1998, Chap. 1) for a parasite called the brain worm. For my present purpose, the important point is to have a clear numerical example of a trait that is selectively disadvantageous within groups but nevertheless increases in frequency in the global population by virtue of a selective advantage between groups.

Knowing all of this, it is possible to imagine an individual deciding to become an A-type based only on its self-interest: “As an A-type, it’s true that I give up a unit of fitness but I also stand a higher chance of being in successful group two. As an S-type, it’s true that I save a unit of fitness but I also stand a higher chance of being in unsuccessful group one. The benefits of group membership outweigh the cost of the behavior; therefore I will become an A-type without caring about anyone else’s fitness.”

This egoistic reasoning process arrives at the correct answer about what evolves because it includes the multi-group population structure in its calculation. It is important to stress that in most real-life situations, an individual who decides to become an A-type is not magically transported into other groups with a probability governed by its newly adopted behavior. It merely loses a unit of fitness compared to behaving as a S-type. The perspective of an omniscient egoist, standing on Mount Olympus and surveying the entire structure of a multi-group population before making its decision, is a heuristic device for the modeler, not a hypothesis that can be taken seriously for humans, much less nonhuman species (Sober, 1998).

In its Olympian form, the concept of self-interest can account for anything that evolves by natural selection, a virtue shared by multilevel selection theory. However, it must never be forgotten that we are trying to answer a specific question – when do groups evolve into adaptive units? Multilevel selection theory’s other virtue is to address this question by distinguishing levels of selection. The Olympian form of self-interest does not share this virtue. The statement "everything that evolves is a form of self-interest" is no help at all. Furthermore, efforts to address the specific question from the Olympian egoistic perspective do not turn up any new principles but merely recreate the distinctions that already exist within multilevel selection theory.

A good example of this futile exercise is selfish-gene theory, which averages the fitness of genes across all contexts and labels whatever evolves as "selfish" because it replaces that which did not evolve (Dawkins, 1976). By itself, the concept of selfish genes is silent on the question of whether individuals or groups qualify as adaptive units. To address this question, a separate concept of "vehicles" must be invoked, which tells us whether genes evolve by coordinating with other genes in the same organism (in which case the individual becomes the vehicle of selection) or the same group (in which case the group becomes the vehicle of selection). In this fashion, selfish gene theory must become as hierarchical as multilevel selection theory to address the issue at the center of Mandeville's fable (Sober & Wilson, 1998; Wilson & Sober, 1994).

In retrospect, I think that the concept of self-interest in both evolution and economics owes its popularity to at least two factors. First, natural selection and rational choice can both often be represented as an optimization process in which a single variable is maximized, even when it is a very complex function of other variables. Second, self-interest is a highly simplified and intuitive case of optimization, in which an individual juggles costs and benefits to maximize a single utility. Thus, any optimization argument can be made intuitive by employing the metaphor of self-interest. Unfortunately, the intuitive appeal of self-interest is obtained at a cost. The details of what constitutes self-interest must change with the particular optimization model, which means that there can be no single concept of self-interest. In addition, optimization in general is poorly suited for addressing the specific question of when groups evolve into adaptive units. Among-group selection by itself maximizes the relative fitness of groups and within-group selection by itself maximizes the relative fitness of individuals within groups, but the entire process of evolution in multi-group populations is a messy combination of these opposing forces, along with other forces that prevent adaptations from evolving in any sense. Showing how groups evolve into adaptive units requires a consideration of opposing forces, not a process in which a single quantity is being maximized.

Psychological Concepts of Self-Interest

All of the preceding concepts of self-interest are based on how individuals act, regardless of how they think or feel. Another set of concepts is based on how

Table 1. Psychological and Evolutionary Self-Interest are Defined on the Basis of Different Criteria, Giving Rise to Four Combinations.

	Evolutionary	
	Selfishness	Altruism
Psychological		
Selfishness		
Altruism		

individuals think and feel, regardless of how they act. As I mentioned previously, one way to relate these two sets of concepts to each other from an evolutionary perspective is with the distinction between proximate and ultimate causation. Adaptive behaviors exist because they cause organisms to survive and reproduce (ultimate causation) but also because of a set of psychological mechanisms that cause the organisms to exhibit the behavior (proximate causation). The question of whether the proximate psychological mechanism counts as self-interested is potentially independent of whether the behavior counts as self-interested in terms of its effects on fitness, which means that all four cells in Table 1 can potentially be occupied (from Sober & Wilson, 1998, p. 204). For example, one individual might be motivated to help others (evolutionarily altruistic) because she values the welfare of others as an end in itself (psychologically altruistic). Such an individual would occupy the lower right-hand cell of Table 1. Another individual might be motivated to help others (evolutionarily altruistic) but only as a means to making herself feel good (psychologically selfish). Such an individual would occupy the upper right-hand cell of Table 1.

Psychological egoism is the claim that all psychological mechanisms that drive behavior count as self-interested, in which case the bottom row in Table 1 would be empty. As with the evolutionary concept of self-interest, we need to ask two questions: Is the claim of psychological egoism true, and does it shed any light on our specific question of when groups evolve into adaptive units?

The first question is evaluated in detail in part II of Sober and Wilson (1998). To make a long story short, the claim cannot be refuted logically but is unlikely to be true based on evolutionary principles. Any behavior, no matter how other-oriented, can theoretically be caused by egoistic proximate mechanisms, but these mechanisms are often like inefficient and error-prone Rube Goldberg devices compared to simpler and more robust non-egoistic mechanisms. After all, if behaviors have been selected to be other-oriented, what simpler way to motivate them than by other-oriented proximate mechanisms? The empirical evidence for exclusive psychological egoism (bottom row of Table 1 empty) is nonexistent. On the contrary, the most careful research points to a mixture of self- and other-oriented mechanisms and the case for exclusive egoism survives only by postulating ever more remote and difficult to test egoistic hypotheses once the simpler ones have been eliminated (e.g. Batson, 1991).

Even more important, the answer to the second question is "no," regardless of the answer to the first question. Just like the extended evolutionary concepts of self-interest, psychological concepts suffer from being too ambitious. In their drive to explain *everything* as a form of self-interest, they lose their ability to ask more focused questions. Let us grant for the sake of argument that the bottom row of Table 1 is empty. Insofar as psychological egoism is sufficiently flexible to motivate both self- and other-oriented behaviors (both columns of Table 1), it remains as silent as selfish gene theory on the factors that cause groups to evolve into adaptive units (the first column vs. the second column).

I will not dwell further on psychological concepts of self-interest, in part because most proponents of self-interest in evolutionary biology and economics do not dwell on them either. Even though self-interest is sometimes defended as the way that people actually think, this position is usually abandoned in favor of the much looser position that people act "as if" they are self-interested, regardless of how they actually think, returning us to the realm of behavior that we have already examined. The impulse to think of genes and organisms that don't think at all (such as bacteria and plants) as self-interested makes it clear that the concept of self-interest is serving primarily as a heuristic device for the human modeler, not a claim about the mechanisms that cause behavior in the proximate sense.

A NEW THEORETICAL FRAMEWORK FOR THE STUDY OF ADAPTIVE HUMAN SOCIETY

I conclude that evolutionary concepts of self-interest do not provide a framework for explaining how groups evolve into adaptive units. The version based on relative fitness within groups shows that self-interest is the problem, not the solution. The more extended versions explain adaptive groupings only insofar as they overlap with the outcome of among-group selection. The most grandiose versions of self-interest become synonymous with "anything that evolves by natural selection" but this is a weakness rather than a strength when it comes to addressing the specific question at the heart of Mandeville's fable.

On the other hand, multilevel selection theory is sufficiently general as an accounting method for evolutionary change, and sufficiently focused on the question of how groups evolve into adaptive units, to provide a theoretical framework for the study of human society. Group selection became such a heresy in the middle of the 20th century that few people know what needs to be retained and what needs to be rejected in the light of subsequent developments. What needs to be retained is the fundamental insight that group-level adaptation requires group-level selection. What needs to be rejected is the notion that among-group selection is always weak compared to within-group selection. Once we appreciate that natural selection is truly a multilevel process, especially in human genetic and cultural evolution, then we can use multilevel selection theory to identify the factors that cause human groups to function well or poorly as adaptive units.

It is beyond the scope of this essay to outline how evolutionary theory in general and multilevel selection theory in particular provide a new theoretical foundation for the study of human society. Fortunately, it is also unnecessary, because a growing number of economists, political theorists, anthropologists, psychologists, and evolutionary biologists are already at work building the foundation (e.g. Boehm, 1999; Bowles & Gintis, 1998; Fehr & Fischbacher, 2002; Henrich, 2004; Hodgson, 1993, 2001; Ofek, 2001; Rubin, 2002; Sober & Wilson, 1998; Wilson, 2002). An excellent example is provided by the recent edited volume *Genetic and Cultural Evolution of Cooperation* (Hammerstein, 2003), which employs the same multilevel selection framework to study the evolution of human society and the evolution of single organisms, with economists playing a key role in the integration.

This essay contributes to the effort by showing that the concept of self-interest does not provide an alternative. My parting generality is that *adaptive human groups must be explained in terms of among-group processes* – in the ancient past, resulting in innate psychological mechanisms that evolved by genetic multilevel selection, and the more recent past, resulting in products of cultural multilevel selection. Switching from individual self-interest to among-group processes as our explanatory framework is a paradigmatic change, even before we consider the details. The new fable of the bees leads in a different direction than Mandeville ever imagined.

NOTE

1. An historically astute reviewer pointed out that Smith did not agree with Mandeville in his *Theory of Moral Sentiments* (III. 1.3). However, Smith's theory of moral sentiments was itself neglected in the development of the modern concept of self-interest. Many of the mechanisms discussed in Smith's book could have emerged from group-level adaptations in terms of multilevel selection theory. They are self-organizing but not self-interested in the sense of evolving by within-group selection.

REFERENCES

Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Lawrence Erlbaum.

Boehm, C. (1993). Egalitarian society and reverse dominance hierarchy. *Current Anthropology*, *34*, 227–254.

Boehm, C. (1999). *Hierarchy in the forest*. Cambridge, MA: Harvard University Press.

Bowles, S., & Gintis, H. (1998). Is equality passe? *Boston Review*, *23*(6), 4–26.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Camazine, S., Deneubourg, J.-L., Franks, N. R., Sneyd, J., Theraula, G., & Bonabeau E. (2001). *Self-organization in biological systems*. Princeton: Princeton University Press.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

Fehr, E., Fischbacher, U., & Gachter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of human social norms. *Human Nature*, *13*, 1–25.

Fehr, E., & Gachter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137–140.

Grafen, A. (1984). Natural selection, kin selection and group selection. In: J. Krebs & N. Davies (Eds), *Behavioural Ecology: An Evolutionary Approach* (pp. 62–84). Oxford: Blackwell Scientific Publications.

Hamilton, W. D. (1975). Innate social aptitudes in man: An approach from evolutionary genetics. In: R. Fox (Ed.), *Biosocial Anthropology* (pp. 133–155). New York: Academic Press.

Hammerstein, P. (2003). *Genetic and cultural evolution of cooperation*. Cambridge, MA: MIT Press.

Heckathorn, D. D. (1990). Collective sanctions and compliance norms: A formal theory of group-mediated social control. *American Sociological Review*, *55*, 366–384.

Heckathorn, D. D. (1993). Collective action and group heterogeneity: Voluntary provision vs. selective incentives. *American Sociological Review*, *58*, 329–350.

Henrich, J. (2004). Cultural group selection, coevolutionary processes, and large-scale cooperation. *Journal of Economic Behavior and Organization*, *53*, 3–55.

Hodgson, G. M. (1993). *Economics and evolution*. Cambridge, UK: Polity Press.

Hodgson, G. M. (2001). *How economics forgot history*. New Brunswick, NJ: Routledge.

Maynard Smith, J., & Szathmary, E. (1999). *The origins of life: From the birth of life to the origin of language*. Oxford: Oxford University Press.

Michod, R. (1999). Individuality, immortality, and sex. In: L. Keller (Ed.), *Levels of Selection in Evolution* (pp. 53–74). Princeton, NJ: Princeton University Press.

Nunney, L. (1985). Group selection, altruism, and structured-deme models. *Am. Nat.*, *126*, 212–230.

Ofek, H. (2001). *Second nature: Economics origins of human evolution*. Cambridge, UK: Cambridge University Press.

Richerson, P. J., & Boyd, R. (in press). *The nature of cultures*. Chicago, IL: University of Chicago Press.

Rissing, S. W., Pollock, G. B. et al. (1989). Foraging specialization without relatedness or dominance among co-founding ant queens. *Nature*, *338*, 420–422.

Rubin, P. (2002). *Darwinian politics: The evolutionary origin of freedom*. New Brunswick, NJ: Rutgers University Press.

Seeley, T. (1995). *The wisdom of the hive*. Cambridge, MA: Harvard University Press.

Seeley, T. D., & Buhrman, S. C. (1999). Group decision making in swarms of honey bees. *Behavioral Ecology and Sociobiology*, *45*, 19–31.

Sober, E. (1998). Three differences between evolution and deliberation. In: P. Danielson (Ed.), *Modeling Rationality, Morality, and Evolution* (pp. 408–422). Oxford: Oxford University Press.

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Wilson, D. S. (1990). Weak altruism, strong group selection. *Oikos*, *59*, 135–140.

Wilson, D. S. (1998). Hunting, sharing and multilevel selection: The tolerated theft model revisited. *Current Anthropology*, *39*, 73–97.

Wilson, D. S. (1999). A critique of R. D. Alexander's views on group selection. *Biology and Philosophy*, *14*, 431–449.

Wilson, D. S. (2002). *Darwin's Cathedral: Evolution, religion and the nature of society*. Chicago: University of Chicago Press.

Wilson, D. S. (2004). What is wrong with absolute individual fitness? *Trends in Ecology and Evolution*, *19*, 245–248.

Wilson, D. S., & Kniffin, K. M. (1999). Multilevel selection and the social transmission of behavior. *Human Nature*, *10*, 291–310.

Wilson, D. S., & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences*, *17*, 585–654.

Uncited reference

References cited in the text must appear in the reference list; conversely, each entry in the reference list must be cited in the text... The author must make certain that each source referenced appears in both places and that the text citation and reference list entry are identical in spelling and year.

Fehr et al. (2002).