15 Imitation and Rationality

Robert Sugden

15.1 Introduction

Conventional economic theory depends heavily on assumptions about the rationality of economic agents. In this chapter, I appraise a theoretical strategy that has been offered as a justification of those assumptions. This strategy adapts Richard Dawkins's (1976, pp. 203–215) idea that human behavior is governed by "memes" that are transmitted from brain to brain through processes of imitation. It treats the rationality of individual agents, not as a property that is intrinsic to human psychology, but as one that emerges through the mutual adaptation of behavior among individuals who have certain tendencies to imitate one another. I argue that this strategy, in the form in which it has so far been used in economics, fails.

15.2 The Evolutionary Turn in Economic Theory

Economic theory has always been vulnerable to the criticism that human beings are not naturally rational in the ways that the theory assumes. Yet, for years, most economists brushed aside such criticisms, asserting that theories based on rationality assumptions generated successful predictions across a wide range of human behavior. There was an element of bluff in that response. Over the past two decades, this bluff has been called by the development of experimental tests of economic theories of decision making. These have revealed many ways in which human decision making deviates systematically from the predictions of conventional rational-choice theories.[1] One effect of these developments has been to prompt economists who favor rational-choice theories to seek reasons why such theories might predict well in the situations in which they are customarily applied, but not in the experimental environments in which they fail.

1. See Camerer (1995) for a survey of this evidence.

In the face of these concerns about the validity of rationality assumptions, many economists have been impressed by the apparent success of rational-choice models of animal behavior in biology. A body of work in theoretical biology, pioneered by John Maynard Smith and G. R. Price (1973), has modeled animal behavior as if it was the solution to the problem of maximizing each animal's reproductive success. Many of the situations studied are remarkably similar to decision problems analyzed in economics. For example, a bird that needs to find food each day for its young may have a range of alternative foraging strategies. If some areas in which food can be collected are richer in food while others are closer to the nest, the selection of the most fitness-enhancing strategy is a classic economic problem of optimization. Other problems animals face, such as when to escalate a conflict and when to back down, are analogous to strategic interaction games. In many such cases, animal behavior can be explained by assuming that each animal acts in the way that maximizes its own reproductive success, given the environment in which it is operating and the behavior that can be expected of other animals. The mechanism that induces these forms of maximizing behavior is natural selection.

What has all this to do with economics? In most of the situations in which economists use rational-choice theory, it is not credible to suppose that the rationality that is attributed to human behavior is a direct product of *biological* natural selection. Biologically, we human beings are adapted to respond to the environment that our ancestors faced in the distant past; but economics assumes that our behavior is a rational response to the problems we *now* face. For economists, evolutionary biology is attractive, not as a theory of economic behavior, but as a demonstration that apparent rationality can be the result of blind processes of selection. The thought is that the rationality of human actors, as represented in economic theory, might be the product of some process of cultural selection or trial-and-error learning, *analogous with* natural selection in biology.

If that could be shown, economists would be able to argue that rational-choice theory applies only in situations in which the relevant selection mechanisms are active. That might make it possible to define the domain of the theory in such a way as to include many forms of behavior in markets while excluding many laboratory experiments.[2] But if the use of evo-

2. Some leading experimental economists have used this line of reasoning to argue that economic theory is intended to predict the behavior in markets of *experienced* traders. Experimental tests of economic theory are accepted as valid only if subjects have been given adequate experience in the tasks they perform (see e.g., Plott, 1996).

lutionary models in economics is to be justified, we need to know which mechanisms in the world of human economic behavior are supposed to be analogous with which biological mechanisms. And we need to be convinced that there are the right kinds of isomorphism between the two sets of mechanisms. Economists have been much more ready to use evolutionary models than to consider, except at the most superficial level, what makes these models valid as representations of real human behavior.[3] Insofar as this issue has been addressed at all, one of the more common ways of justifying evolutionary modeling in economics has been to argue that the mechanism of selection is one of imitation.

15.3 Imitation as a Selection Mechanism

In this chapter I am concerned with a particular version of the argument that rationality is selected through imitation: the argument presented by Ken Binmore (1994, 1998). Binmore's approach is inspired by Dawkins's (1976) concept of a meme. The idea rests on a simple but radical analogy with biological natural selection. In biology, the apparent rationality of animal behavior is as if it was directed toward the objective of maximizing reproductive success, measured by the replication of genes. Why is the replication of genes the objective? Because an animal's behavior is governed by the genes it carries, and because natural selection *is* the differential replication of genes. So (the argument goes), if social evolution is to be understood through the analogy of biological evolution, we need to find what, in the social selection of behavior, plays the role of a gene. If selection takes place through imitation, what we are looking for is something that is transmitted by imitation—a meme. (The word was coined by Dawkins to suggest both imitation—the Greek root is *mimeomai*, to imitate—and an analogy with gene.) The closest Dawkins comes to a description of what a meme is is this:

> Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation. (Dawkins, 1976, p. 206)

In the analogy of biological natural selection, we should expect to find that social selection favors those forms of behavior that maximize the

3. I substantiate this criticism in Sugden (2001).

replication of their own memes. Binmore takes up Dawkins's idea and uses it to attempt to explain the supposed tendency for human beings to act as if they had consistent preferences of the kind assumed by rational-choice theory.

Binmore's argument is part of a massive treatise that offers a social evolutionary explanation of certain normative principles of justice. Binmore claims to offer a naturalistic account of why, as a matter of sociological fact, these principles are treated as having normative force in a wide range of human societies. The argument depends crucially on modes of analysis that are taken from rational-choice theory. It would be inconsistent for Binmore, as a philosophical naturalist, to appeal to the supposed normative force of rationality principles. Instead, he begins by declaring that if it is valid to model people as maximizers—as his own theory will do—this can only be because "evolutionary forces, biological, social, and economic, [are] responsible for getting things maximized" (Binmore, 1994, p. 20).

He then appeals to Dawkins's analysis of memes, arguing that there is a social evolutionary process that eliminates "inferior" memes:

> In this story, people are reduced to ciphers. Their role is simply to carry memes in their heads, rather as we carry the virus for the common cold in the winter. However, to an observer, it will seem as though the infected agent is acting in his own self-interest, provided that the notion of self-interest is interpreted as being *whatever makes the bearer of a meme a locus for replication of the meme to other heads.* (Binmore, 1994, p. 20)

Thus, according to Binmore, there is a tendency for social selection to favor behavior that is as if it was motivated by consistent preferences:

> the practical reasons for thinking consistency an important characteristic of a decision-maker cannot be lightly rejected. People who are inconsistent will necessarily sometimes be wrong and hence will be at a disadvantage compared to those who are always right. And evolution will not be kind to memes that inhibit their own replication. (Binmore, 1994, p. 27)

Notice that this argument does not presuppose any criterion of successful behavior which then directs imitation. The primitive concept in the theory is imitation itself. Whatever tends to be imitated thereby has a tendency to be replicated. When the process of selection has run its course, human behavior will be as if it was motivated by a criterion of success. But in this state of affairs, "successful" actions are not imitated because they are successful; *being imitated is what they are successful at.*

Binmore's meme-based argument should not be confused with other, less radical claims about the connections between imitation and rationality. In particular, it must be distinguished from claims about imitation that presuppose particular criteria of success. For example, if we assume that individuals are motivated by particular preferences, there may be circumstances in which imitation is effective as a rule of thumb for satisfying those preferences in an uncertain environment. (Consider the rule sometimes recommended to tourists, of choosing restaurants that appear to be well patronized by local people.) Imitation may also appear in evolutionary models as the mechanism by which selection works, given some criterion of success that is independent of imitation. For example, many applications of evolutionary game theory in economics presuppose that utility indices can be attached to the outcomes of games that are played recurrently in a population. Evolutionary selection in such models is simply a tendency for behavior to gravitate toward those strategies that maximize expected utility. Imitation—or more precisely, a tendency disproportionately to imitate the behavior of individuals who have been seen to be relatively successful in gaining utility—is often suggested as a mechanism that might lie behind this gravitation.[4]

Social evolutionary models that depend on assumed criteria of success have many applications, but no such model can provide a justification for the assumptions of rational-choice theory. The reason is simple. The most fundamental assumptions of rational-choice theory are *equivalent to* the assumption that there is a one-dimensional criterion of success for human action (that is, expected utility). In the absence of arguments from rationality, it is not clear what grounds we have for supposing there to be any such criterion. On the face of it, any human action can be described on many different dimensions. What makes all these different dimensions commensurable? The answer given by rational-choice theory is that commensurability is a product of rationality, and that the scale of measurement is that of subjective but rationally consistent preference. If we want an evolutionary *explanation* for the (supposed) fact that human beings are rational in the sense of rational-choice theory, we have to find a criterion of "success" that does not presuppose internally consistent preferences. We then have to show that selection tends to eliminate behavior that fails to maximize success, so defined. That is what Binmore claims to do.

4. Weibull (1995) presents a family of imitation models of this kind.

15.4 The Rational Replicator

According to Dawkins, genes are selfish. In his metaphor, animals are "survival machines" built by genes (1976/1989, p. 21). It is as if genes are the active agents in the biological world, each gene rationally seeking to replicate itself. Similarly, we are asked to think of memes as the active agents in the world of human culture, rationally seeking to replicate themselves in the medium of human minds. As a first step in analyzing Binmore's argument, it is useful to be clear about the sense in which genes and memes can be said to be (as if) rational.

Consider the following abstract model of replication. (Models of this kind are known in mathematical biology as Lotka-Volterra models.) Suppose there are things called *replicators*. At this stage, I do not specify what replicators are, but the concept is intended to encompass both genes and memes. These replicators come in m discrete *types*. There is a *population* made up of very large numbers of replicators of these various types. At any moment in time t, for each type i, there is a *frequency* $p_i[t]$, so that $\Sigma_i p_i[t] = 1$; this represents the proportion of the whole population of replicators that is of type i. The frequency distribution $(p_1[t], \ldots, p_m[t])$ is denoted by $p[t]$. Each replicator is capable of creating copies of itself (but not of other types of replicator). The *replication rate* of any type at any given time is defined as the average rate, per unit of time, at which each replicator of that type is creating copies of itself. (If replicators can "die," deaths are treated as negative copies.) Suppose that the replication rate of each type varies continuously with the distribution of types in the population, but otherwise is independent of time. Then, for each type i, we can define a continuous *replication function* $r_i(.)$ so that the replication rate of type i at time t is $r_i(p[t])$. We have now fully specified the dynamic process by which, starting from any arbitrary $p[0]$, the distribution of types in the population changes over time; this is the process of *replicator dynamics* (P. Taylor and Jonker, 1978).

A *rest point* in this process is a frequency distribution that persists indefinitely. A rest point p^* is *stable* if, starting from any frequency distribution sufficiently close to p^*, the dynamic process converges to p^*. Let us say that type k *survives* at a stable rest point p^* if $p_k^* > 0$. And let us say that type k *maximizes replication* at p if, for all i, $r_k(p) \geq r_i(p)$. It is easy to see that for any stable rest point p^*, any given type k survives at p^* only if it maximizes replication at p^*. (Once the path of $p[t]$ has come close to p^*, any type that maximizes replication will gradually increase its frequency relative to any type that does not.)

So, if the population of replicators is in a state of stable equilibrium, the replicators that survive in that population "act" in ways that maximize their own replication rates. In this special sense, replicators are acting as if they are rational. Thus, if behavior in some animal species is determined by genes, and if genes are replicators in the sense of the model I have set out, then in a state of equilibrium, *genes* are acting as if they are rational. This corresponds with Dawkins's picture of the "selfish gene." Similarly, if human actions in the domain of economics are determined by memes, and if memes are replicators in the sense of the model, then *memes* are acting as if they are rational. But what does the rationality of genes or memes tell us about the rationality of *actions*?

15.5 Replicators and Actors: The Simplest Model

In order to answer this question, we need a model that includes *actors* as well as replicators. In theories of animal behavior, the actors are individual animals; in Binmore's theory, they are individual human beings. The model has to represent both the mechanism by which replicators determine actors' choices among actions, and the mechanism by which the chosen actions determine the rates at which replicators replicate.

I present three alternative models of this causal loop. These models are offered as thought experiments, illustrating theoretical possibilities. I must emphasize that I am not proposing a theory of memes; I am examining claims that other theorists have made.

To keep things as simple as possible, the recurrent problems I consider are games against nature (that is, decision problems that do not involve strategic interaction between individuals). In all my models, I assume a finite set of *consequences*, X. A *decision problem* is a nonempty subset of X. A typical decision problem will be written as $\{x_1, \ldots, x_n\}$. The interpretation is that any actor who faces this problem has to choose one action from a set of n alternative actions; each action leads to a distinct consequence.

Conventional rational-choice theory requires that an individual's choices reveal a preference ordering among possible consequences. The issue to be investigated is whether selection at the level of replicators induces this form of rationality at the level of actors. In a model in which rationality is to have some chance of being induced by imitation, preference must be treated as a property of a population of actors, not as a property of any individual actor. So the idea to be tested is whether through mutual imitation within such a population, all the members of that population converge on a single pattern of choice that can be represented by a preference ordering.

Consider any fixed decision problem $\{x_1, \ldots, x_n\}$ faced recurrently by individual actors in a large population. At any time t, for each consequence x_j, there is a *decision probability* $P_j[t]$; this is the probability that in a randomly selected instance of the decision problem within the population of actors, that consequence will be chosen. Obviously, $\Sigma_j P_j[t] = 1$ at all t. Decision probabilities in the population of actors are assumed to be determined in some way by the relative frequencies of different types of replicators in a *replicator pool*. In turn, the replication rates of the different types of replicators are determined by the consequences that are chosen in the population of actors, and hence by the decision probabilities.

At this level of generality, the modeling framework can be interpreted either in terms of biological natural selection or in terms of imitation. If it is interpreted biologically, the replicators are genes, and each actor's actions are determined by the genes that it carries. If it is interpreted in terms of imitation, the actors are human beings, the replicators are memes, and each individual's actions are determined by the memes that he or she carries.

I begin with the simplest possible model. This can be interpreted as a highly simplified representation of how animal behavior is determined by biological natural selection.[5] It rests on two crucial (and unrealistic) assumptions.

The first assumption is that there is a one-to-one correspondence between genes and actions. The recurrent decision problem is a choice from an opportunity set of n consequences, $\{x_1, \ldots, x_n\}$. One piece of genetic code is responsible for determining which consequence is chosen in this problem, and is responsible for nothing else. There are exactly n alternative versions of this genetic code; each actor carries one and only one of these codes. I call these alternative codes "genes." (Biologists might prefer to call them alternative "alleles" of a single gene.) Each gene is associated with a distinct consequence in the decision problem, in such a way that an actor carrying the gene for some particular consequence invariably chooses the action that leads to that consequence.

The second assumption is that reproduction is asexual. When an actor reproduces, it produces offspring that are genetic copies of itself.

Given these assumptions, we can use the same indices $j = 1, \ldots, n$ for consequences and genes; the jth gene is defined as the gene that programs the choice of consequence x_j. Since each actor carries one and only one gene, and since each gene programs a distinct action, the frequency p_j of the jth gene in the gene pool is identically equal to the decision probability P_j for the consequence x_j. This requires that the dynamics of replication within the gene pool be reflected exactly in the dynamics of changes in decision probabilities.

5. I draw on Binmore's (1992, pp. 414–422) explanation of replicator dynamics as a representation of the life cycle of an imaginary (and biologically peculiar) species.

Now consider how decision probabilities induce changes in the gene pool. For each consequence x_j, we can define a measure $R(x_j)$ of the *reproductive success* conferred on the actor by that consequence. Reproductive success is to be understood in terms of expected numbers of offspring. More precisely, averaging over all those actors that are genetically programmed to choose x_j, $R(x_j)$ is the rate at which these actors are producing genetic copies of themselves (with deaths counting as negative copies). I assume that for each x_j, the value of $R(x_j)$ is constant over time and independent of other consequences in the opportunity set.

In this model, reproductive success for an actor corresponds with replication of the gene that the actor carries, since reproduction is the creation of exact genetic copies. Thus, the replication rate of the jth gene is equal to $R(x_j)$. This rate is constant over time and is independent of the frequencies of the different types of gene in the gene pool. Clearly, genes with higher replication rates will increase in relative frequency in the gene pool at the expense of those with lower rates. Thus, if one consequence, say x_k, leads to strictly greater reproductive success than every other consequence, the relative frequency of the corresponding gene will increase continuously. So the dynamics of the gene pool will converge to a stable rest point at which only the kth gene survives. Correspondingly, the dynamics of decision probabilities will converge to a stable rest point at which x_k is chosen with a probability of 1. (If two or more consequences have equal reproductive success, and greater reproductive success than all other consequences, the sum of the decision probabilities for the consequences with greatest reproductive success will converge to 1.)

The implication is that after natural selection has run its course, actors behave as if they are rational in the sense of rational-choice theory. For each consequence x_j, there is an index $R(x_j)$ that is independent of the particular decision problem in which that consequence is located. In the long run, actors behave as if they are maximizing the value of this index. Thus, this index plays the role of a utility index in rational-choice theory; the ranking of consequences generated by this index plays the role of a preference ordering.

This model shows one way in which imitation could conceivably induce rationality. *If* the relationship between memes and actions in the real world

was isomorphic with the relationship between genes and actions in this model, *then* selection operating on memes would induce rationality on the part of actors.

Here is a stylized example of how this could come about. Consider an artisan trade before the industrial revolution. The unit of organization is the workshop, owned by a master craftsman. Young men enter the trade by being apprenticed to a master, from whom they learn the skills of the craft; they then work on their own account. Let $\{x_1, \ldots, x_n\}$ be a set of alternative techniques. Suppose that each craftsman uses just one of these techniques; his apprentices learn this technique by imitation and then use it themselves. In this model, the master craftsmen are actors and the techniques are actions. "Reproductive success" for a master is measured by the number of his former apprentices who set up as masters. For each technique x_j, we can define an index $R(x_j)$ that measures the reproductive success (as just defined) of masters who use that technique, and hence also the replication rate for that technique. In the long run, the behavior of masters in choosing among techniques will be as if they are trying to maximize the value of the function $R(.)$. This pattern of behavior is rational in the sense of rational-choice theory (it maximizes *something*), even though it is not necessarily rational in the sense of maximizing each master's profits.

So this first model offers some support for the hypothesis that imitation induces rationality. However, the model represents a very simple relationship between replicators and actions. The relationship is one-to-one at both sides of the causal loop. Each replicator is the cause of one and only one action, and each action is capable of creating copies only of the replicator that causes it. In a model with this structure, "rationality" in the domain of replicators *does* induce rationality in the domain of actions. But what if the relationship between replicators and actions is not quite so simple?

15.6 Replicators and Actors: Sexual Reproduction

My second model, like the first, is based on biology. The only change I make to the model presented in section 15.5 is to introduce sexual reproduction.

In a sexually reproducing (diploid) species, each individual's genetic inheritance comes from two parents. Because of this fact, we must distinguish between *genes* (understood as the units of genetic material that are transmitted through reproduction) and *genotypes* (that is, alternative bundles of genes that an individual can inherit). To keep the model as simple as possible, I consider only one genetic "locus." Each actor inherits two genes, one from each parent. On the assumption that mating is random, each of these two genes can be thought of as resulting from a random draw from the same gene pool. The resulting *pair* of genes is the actor's genotype. I assume that each actor's behavior is uniquely determined by its genotype. As in the first model, each consequence has an associated measure of reproductive success, interpreted as the expected number of offspring for an actor who experiences it. However, to each of its offspring, each parent passes on only *one* of its pair of genes; which of the two is passed on is determined by a random process. In this model, selection does not necessarily eliminate behavior that fails to maximize reproductive success. The following example shows why.

Suppose there are just two genes: A and a. This gives three alternative genotypes: AA, Aa, and aa. (AA and aa are *homozygous*; Aa is *heterozygous*.) Consider the decision problem $\{x_1, x_2, x_3\}$. Suppose that actors with the AA genotype choose x_1, that Aa actors choose x_2, and that aa actors choose x_3. The decision probabilities for actions are uniquely determined by the relative frequencies of genes in the gene pool, but the link between the two probability distributions is more complicated than in the first model. Specifically, let q be the proportion of A genes in the gene pool. Then the decision probabilities for consequences x_1, x_2, x_3 are given by $P_1 = q^2$, $P_2 = 2q(1 - q)$, and $P_3 = (1 - q)^2$.

Now suppose that $R(x_2) > R(x_1) > R(x_3)$. Recall that $R(x_j)$ measures the contribution made by x_j to the reproductive success *of the actor who chooses it*. So the assumption is that x_2 is the consequence that maximizes reproductive success. But actors who choose x_2 carry the genotype Aa. When they reproduce, they create copies of *both* genes. Thus, both genes will survive in the gene pool and, as a consequence of this, all three genotypes will persist, and all three consequences will be chosen with positive probability. It is even possible that the consequence with the largest decision probability is not the one with the greatest reproductive success. For example, suppose that $R(x_1)/R(x_2) = 0.9$ and $R(x_3)/R(x_2) = 0.2$. It turns out that in equilibrium, $q = 0.89$,[6] which implies the decision probabilities $P_1 = 0.79$, $P_2 = 0.20$, $P_3 = 0.01$.

The implication of this model is that biological natural selection does not necessarily favor *actions* that maximize the reproductive success *of actors*. The evolution of decision probabilities can gravitate toward a stable rest point at which each of several actions is chosen with positive probability, even though these actions have very different degrees of reproductive

6. The general result is that $q/(1 - q) = [1 - R(x_3)/R(x_2)]/[1 - R(x_1)/R(x_2)]$. This is derived from the condition that copies of the genes A and a are made in the ratio $q : (1 - q)$, thus conserving the ratio in the gene pool.

success. The less successful actions survive, not because of *their* propensity to replicate the genes that cause them to be chosen, but because those genes are also replicated by other, more successful actions. This biological mechanism accounts for the genetic transmission of certain diseases, such as sickle-cell anemia. In this type of case, the aa genotype leads to the disease, but the Aa genotype gives its carriers some gain in fitness relative to those who carry AA.

This paradoxical result is entirely consistent with the idea of the "rational replicator." More precisely, the equilibrium I have described for the two-gene model is a stable rest point in the dynamics of the gene pool, at which both genes survive. At this rest point, the two genes have equal rates of replication. Thus, it is as if the surviving *genes* are maximizing their own replication. But rationality (in this sense) at the level of genes does not induce rationality at the level of actors. The source of the paradox is that the choice of an action is determined, not by a single gene, but by a combination of genes, and that each action has a tendency to replicate each of the genes in the combination that leads to its being chosen. Why should the same not be true of memes?

Seen in relation to the theoretical strategy of explaining economic rationality as the result of selection at the level of memes, this result is discouraging. That strategy treats both the concept of a meme and the process by which memes replicate as black boxes. All that is observed is the behavior of actors in response to decision problems. The objective of the theory is to explain that behavior. The theory depends on the hypothesis that meme selection mechanisms *in general* favor behavior that is rational at the level of actors. But it seems that that hypothesis is false.

15.7 Replicators and Actors: Mutual Imitation

In the second model—the model with sexual reproduction—the hypothesis that selection always induces rationality fails because the distribution of replicators in the replicator pool does not map in a straightforward way onto decision probabilities. My final model shows the effects of disrupting the other side of the simple causal loop of the model presented in section 15.5.

The model I now present is specifically intended to represent imitation among human beings.[7] The population of actors is taken to be fixed; actors

7. A more general form of the model discussed in this section, applying to choice under uncertainty and not formulated in the language of memes, is presented by Cubitt and Sugden (1998).

do not reproduce or die. As in the first model, there is a one-to-one correspondence between replicators and consequences. At any given time, each actor carries one and only one meme. Actors face the decision problem recurrently. For each consequence x_j there is a distinct meme such that, if an actor who is currently carrying that meme confronts the decision problem, she chooses x_j. Thus, the frequency distribution of memes in the meme pool corresponds exactly with the distribution of decision probabilities in the population of actors. The difference from the first model concerns the mechanism by which memes replicate. The replication mechanism in the present model is intended to represent a fundamental property of imitation—that imitation involves two actors, the actor who imitates and the actor who is imitated. It works as follows.

At random intervals, ordered pairs of actors drawn at random from the population meet one another. One actor is the *reviewer*, the other the *comparator*. The reviewer compares the consequence that she experiences, say x_j, with the consequence that the comparator experiences, say x_k. This comparison leads to one of two results. Either the reviewer imitates the comparator, or she does not. In terms of memes, either the reviewer comes to carry the meme for x_k, or she continues to carry the meme for x_j. In the former case, the comparator's meme has replicated itself (and the reviewer's original meme has been displaced).

In the biological models presented in sections 15.5 and 15.6, reproductive success is a property of consequences. In those models, each index of reproductive success $R(x_j)$ is a measure of the degree to which the occurrence of x_j produces offspring for the actor who experiences that consequence. Thus $R(x_j)$ also measures the degree to which the occurrence of x_j produces copies of the gene or genes carried by that actor. To find an analogue of $R(.)$ in the present model, we need to consider how the occurrence of particular consequences induces changes in the composition of the meme pool. The crucial feature of this model is that such changes are induced, not by the occurrence of single consequences, but by the occurrence of pairs of consequences.

For any ordered pair of consequences (x_j, x_k), we can define an *imitation probability* $M(x_j, x_k)$. This is the probability that, conditional on a meeting between a reviewer carrying the meme for x_j and a comparator carrying the meme for x_k, the reviewer comes to carry the comparator's meme (that is, the reviewer imitates the comparator's pattern of behavior). Note that because meetings are random, the probability of a meeting (in any given short period of time) between a reviewer carrying the jth meme and a comparator carrying the kth meme is equal to the probability of a meeting between a reviewer carrying the kth meme and a comparator carrying

the jth meme. Thus, averaging over both types of meeting, we can treat $[M(x_k, x_j) - M(x_j, x_k)]/2$ as a measure of *net growth* in the numbers of carriers of the jth meme, per meeting between a carrier of one meme and a carrier of the other. To simplify the notation, I define a function $\varphi(.,.)$ so that $\varphi(x_j, x_k) = [M(x_k, x_j) - M(x_j, x_k)]/2$. Note that, by construction, this function is skew-symmetric; for all j and k, $\varphi(x_k, x_j) = -\varphi(x_j, x_k)$.

Clearly, if the decision problem contains only two consequences, x_1 and x_2, the process of imitation will favor whichever consequence has positive net growth in comparisons between the two of them. If, say, $\varphi(x_1, x_2) > 0$, P_1 will increase continuously at the expense of P_2; the dynamics will lead toward a stable rest point at which $P_1 = 1$. Thus, *in relation to any given binary decision problem*, this model implies that imitation selects behavior that is as if it was governed by a preference relation; that preference relation can be derived from $\varphi(.,.)$ by reading each $\varphi(x_j, x_k) > 0$ as "x_j is preferred to x_k."

However, this does not imply that behavior in decision problems *in general* has the properties postulated by rational-choice theory. The problem is that rational-choice theory requires the preference relation to be transitive. Nothing that I have said so far imposes the corresponding requirement on the process of imitation. For example, suppose there are three consequences x_1, x_2 and x_3 so that $\varphi(x_1, x_2) > 0$, $\varphi(x_2, x_3) > 0$ and $\varphi(x_3, x_1) > 0$. In words, comparisons between x_1 and x_2 induce a net growth in the number of actors choosing x_1; comparisons between x_2 and x_3 induce a net growth in the number of actors choosing x_2; and comparisons between x_3 and x_1 induce a net growth in the number of actors choosing x_3. When we bring together the implications of the model for the three binary decision problems $\{x_1, x_2\}$, $\{x_2, x_3\}$ and $\{x_3, x_1\}$, we find that these implications are not consistent with any preference ordering over the consequences x_1, x_2, x_3.

What happens if the decision problem is $\{x_1, x_2, x_3\}$? In this case, the dynamics of the model typically induce cycles in the values of the decision probabilities. These probabilities change continuously, but never converge to any rest point. From the viewpoint of rational-choice theory, such a pattern of behavior *at the level of actors* is inexplicable. At the level of memes, it might be said, these cycles make perfectly good sense. They reflect the fact that given the assumed properties of the imitation process, the rate of replication for any one meme depends on the relative frequencies of all three memes in the meme pool. But, in relation to the argument of this chapter, that is beside the point. The question at issue is whether selection acting on memes induces rationality at the level of actors—to which the answer must be "not necessarily."

To this, it might be objected that the cyclical pattern of imitation that I have hypothesized is incoherent or pathological, but this objection makes an implicit appeal to a criterion of "success" for actions other than imitation itself. Obviously, if we *presuppose* a one-dimensional measure of success for actions, and interpret imitation as the imitation *of success*, so defined, then cyclical patterns are not coherent. However, such a presupposition is incompatible with the theoretical strategy that I am appraising. It is an essential part of that strategy that no prior measure of success be assumed. So the implications to be drawn from this model are similar to those to be drawn from the model of sexual reproduction. Selection at the level of memes does not necessarily induce rationality at the level of actors.

15.8 Conclusion

Dawkins's original discussion of memes, written as a postscript to a book about natural selection in biology, is a heady mix of brilliant insight, imaginative speculation, and scientific hubris. (The hubris comes in the scarcely veiled suggestion that the investigation of memes is an intellectual greenfield site, ripe for development by biologists. There is no mention of the possibility that disciplines such as linguistics, art history, or economic history might be the various forms that the study of memes already takes.) The crucial thought is that within human populations, ways of thinking and patterns of decision making are not selected for the degree to which they serve the interests of human beings; they are selected for the degree to which they induce whatever conditions promote their own replication. On first reading, this is a startling claim, but I am convinced that it expresses an important truth. Nevertheless, social scientists need to be careful not to be carried away by Dawkins's rhetoric.

For mathematical theorists, I think, one of the seductive features of Dawkins's treatment of memes is its a priori character. It appears to be deriving significant conclusions about cultural transmission without any messy investigation of facts. Instead, it points to apparent analogies between cultural transmission and certain biological mechanisms that evolutionary game-theoretical models have already helped us to understand. The temptation is to think that we can arrive at a similar understanding of human imitation merely by importing those models into social science.

The truth is that biology is much more than evolutionary game theory. In particular, biological theories of natural selection depend on biologists' empirically grounded understanding of what genes are and the mechanisms by which they replicate. Without this kind of understanding, natural

selection would not be the theory it is, but merely the tautology that in any pool of replicators, those replicators that are more successful at replicating will increase in frequency relative to those that are less successful. The "theory" of memes, as used in the arguments I have been appraising, *is* only that tautology. What is missing is an understanding of what memes actually are and how they in fact replicate. And that understanding is not possible without an investigation of the facts of cultural transmission.

Whether human decision-making behavior satisfies the rationality postulates of conventional choice theory is an empirical question. If it does, any explanation of that fact must depend on empirical propositions about how the world really is. Trying to find an explanation by manipulating tautologies about replicators is to attempt what is logically impossible.[8]

Acknowledgments

Many of the ideas in this paper were developed in collaboration with Robin Cubitt. I am grateful for comments from Mark Greenberg and Paul Seabright. My work has been supported by the Leverhulme Trust.

8. See comments on this chapter by Seabright (vol. 2, ch. 19.10, p. 398), and Greenberg (vol. 2, ch. 19.11, p. 402). ED.