12

The Unification of the Behavioral Sciences

> Each discipline of the social sciences rules comfortably within its own chosen domain. . . so long as it stays largely oblivious of the others.
>
> Edward O. Wilson

> The combined assumptions of maximizing behavior, market equilibrium, and stable preferences, used relentlessly and unflinchingly, form the heart of the economic approach
>
> Gary Becker

> While scientific work in anthropology, and sociology and political science will become increasingly indistinguishable from economics, economists will reciprocally have to become aware of how constraining has been their tunnel vision about the nature of man and social interaction.
>
> Jack Hirshleifer

The behavioral sciences include economics, anthropology, sociology, psychology, and political science, as well as biology insofar as it deals with animal and human behavior. These disciplines have distinct research foci, but they include four conflicting models of decision making and strategic interaction, as determined by what is taught in the graduate curriculum and what is accepted in journal articles without reviewer objection. The four are the psychological, the sociological, the biological, and the economic.

These four models are not only different, which is to be expected given their distinct explanatory aims, but are also *incompatible*. That is, each makes assertions concerning choice behavior that are denied by the others. This means, of course, that at least three of the four are certainly incorrect, and I will argue that in fact all four are flawed but can be modified to produce a unified framework for modeling choice and strategic interaction for all of the behavioral sciences. Such a framework would then be enriched in different ways to meet the particular needs of each discipline.

In the past, cross-disciplinary incoherence was tolerated because distinct disciplines dealt largely with distinct phenomena. Economics dealt with market exchange. Sociology dealt with stratification and social deviance. Psychology dealt with brain functioning. Biology, failing to follow up on Darwin's insightful monograph on human emotions (Darwin 1998),

avoided dealing with human behavior altogether. In recent years, however, the value of transdisciplinary research in addressing questions of social theory has become clear, and sociobiology has become a major arena of scientific research. Moreover, contemporary social policy involves issues that fall squarely in the interstices of the behavioral disciplines, including substance abuse, crime, corruption, tax compliance, social inequality, poverty, discrimination, and the cultural foundations of market economies. Incoherence is now an impediment to progress.

My framework for unification includes five conceptual units: (a) gene-culture coevolution; (b) the sociopsychological theory of norms; (c) game theory, (d) the rational actor model; and (e) complexity theory. Gene-culture coevolution comes from the biological theory of social organization (sociobiology) and is foundational because *H. sapiens* is an evolved, highly social, biological species. The sociopsychological theory of norms includes fundamental insights from sociology and social psychology that apply to all forms of human social organization, from hunter-gatherer to advanced technological societies. These societies are the product of gene-culture coevolution but have *emergent properties* (§8.8), including social norms and their psychological correlates/prerequisites, that cannot be derived analytically from the component parts of the system—in this case the interacting agents (Morowitz 2002).

Game theory includes four related disciplines: classical, behavioral, epistemic, and evolutionary game theory, the first three of which have been developed in this book. The fourth, evolutionary game theory, is a macro-level analytical apparatus allowing biological and cultural evolution to be mathematically modeled.

The rational actor model (§1.1, 1.5) is the most important analytical construct in the behavioral sciences operating at the level of the individual. While gene-culture coevolutionary theory is a form of ultimate explanation that does not predict, the rational actor model provides a proximate description of behavior that can be tested in the laboratory and in real life and is the basis of the explanatory success of economic theory. Classical, epistemic, and behavioral game theory make no sense without the rational actor model, and behavioral disciplines, such as anthropology and sociology, as well as social and cognitive psychology, that have abandoned this model have fallen into theoretical disarray.

Behavioral economists and psychologists have taken aim at the rational actor model in the belief that experimental results contradict rationality.

Showing that this view is wrong has been a constant theme of this book. The behaviorists' error is partly due to their having borrowed a flawed conception of rationality from classical game theory, partly due to their interpreting the rational actor model too narrowly, and partly due to an exuberant but unjustified irreverence for received wisdom.

Complexity theory is needed because human society is a complex adaptive system with *emergent properties* that cannot now be, and perhaps never will be, fully explained starting with more basic units of analysis. The hypothetico-deductive methods of game theory and the rational actor model, and even gene-culture coevolutionary theory, must therefore be complemented by the work of behavioral scientists who deal with society in more macrolevel, interpretive terms, and develop insightful schemas that shed light where analytical models cannot penetrate. Anthropological and historical studies fall into this category, as well as macroeconomic policy and comparative economic systems. Agent-based modeling of complex dynamical systems is also useful in dealing with emergent properties of complex adaptive systems.

The above principles are not meant to revolutionize research in any discipline. Indeed, they build on existing strengths, and they imply change only in the areas of overlap among disciplines. For instance, a psychologist working on visual processing, or an economist working on futures markets, or an anthropologist documenting food-sharing practices, or a sociologist gauging the effect of dual parenting on children's educational attainment might gain little from knowing that a unified model of decision making underlies all the behavioral disciplines. On the other hand, a unified model of human choice and strategic interaction might foster innovations that come to pervade the discipline, even in these relatively hermetically sealed areas.

12.1 Gene-Culture Coevolution: The Biological Model

The centrality of culture and complex social organization to the evolutionary success of *H. sapiens* implies that individual fitness in humans depends on the structure of social life. Since culture is limited and facilitated by human genetic propensities, it follows that human cognitive, affective, and moral capacities are the product of an evolutionary dynamic involving the interaction of genes and culture. This dynamic is known as *gene-culture coevolution* (Cavalli-Sforza and Feldman 1982; Boyd and Richerson 1985; Dunbar 1993; Richerson and Boyd 2004). This coevolutionary

process has endowed us with preferences that go beyond the self-regarding concerns emphasized in traditional economic and biological theory and embrace a social epistemology facilitating the sharing of intentionality across minds, as well as such non-self-regarding values as a taste for cooperation, fairness, and retribution, the capacity to empathize, and the ability to value honesty, hard work, piety, toleration of diversity, and loyalty to one's reference group.

Gene-culture coevolution is the application of *sociobiology*, the general theory of the social organization of biological species, to species that transmit culture without informational loss across generations. An intermediate category is *niche construction*, which applies to species that transform their natural environment to facilitate social interaction and collective behavior (Odling-Smee, Laland, and Feldman 2003).

The genome encodes information that is used both to construct a new organism and to endow it with instructions for transforming sensory inputs into decision outputs. Because learning is costly and error-prone, efficient information transmission ensures that the genome encodes all aspects of the organism's environment that are constant or that change only slowly through time and space. By contrast, environmental conditions that vary rapidly can be dealt with by providing the organism with the capacity to *learn*.

There is an intermediate case, however, that is efficiently handled neither by genetic encoding nor by learning. When environmental conditions are positively but imperfectly correlated across generations, each generation acquires valuable information through learning that it cannot transmit genetically to the succeeding generation because such information is not encoded in the germ line. In the context of such environments, there is a fitness benefit to the transmission of *epigenetic* information concerning the current state of the environment.[1] Such epigenetic information is quite common (Jablonka and Lamb 1995) but achieves its highest and most flexible form in *cultural transmission* in humans and to a considerably lesser extent in other primates (Bonner 1984; Richerson and Boyd 1998). Cultural transmission takes the form of vertical (parents to children), horizontal (peer to peer), and oblique (elder to younger), as in Cavalli-Sforza and Feldman (1981), prestige (higher status influencing lower status), as in Henrich and Gil-White (2001), popularity-related as in Newman, Barabasi, and Watts

[1] An epigenetic mechanism is any nongenetic intergenerational information transmission mechanism, such a cultural transmission in humans.

(2006), and even random population-dynamic transmission, as in Shennan (1997) and Skibo and Bentley (2003).

The parallel between cultural and biological evolution goes back to Huxley (1955), Popper (1979), and James (1880)—see Mesoudi, Whiten, and Laland (2006) for details. The idea of treating culture as a form of epigenetic transmission was pioneered by Richard Dawkins, who coined the term "meme" in *The Selfish Gene* (1976) to represent an integral unit of information that could be transmitted phenotypically. There quickly followed several major contributions to a biological approach to culture, all based on the notion that culture, like genes, could evolve through replication (intergenerational transmission), mutation, and selection.

Cultural elements reproduce themselves from brain to brain and across time, mutate, and are subject to selection according to their effects on the fitness of their carriers (Parsons 1964; Cavalli-Sforza and Feldman 1982). Moreover, there are strong interactions between genetic and epigenetic elements in human evolution, ranging from basic physiology (e.g., transformation of the organs of speech with the evolution of language) to sophisticated social emotions, including empathy, shame, guilt, and revenge seeking (Zajonc 1980, 1984).

Because of their common informational and evolutionary character, there are strong parallels between genetic and cultural modeling (Mesoudi, Whiten, and Laland 2006). Like biological transmission, cultural transmission occurs from parents to offspring, and like cultural transmission, which occurs horizontally between unrelated individuals, in microbes and many plant species, genes are regularly transferred across lineage boundaries (Jablonka and Lamb 1995; Rivera and Lake 2004; Abbott et. al 2003). Moreover, anthropologists reconstruct the history of social groups by analyzing homologous and analogous cultural traits, much as biologists reconstruct the evolution of species by the analysis of shared characters and homologous DNA (Mace and Pagel 1994). Indeed, the same computer programs developed by biological systematists are used by cultural anthropologists (Holden 2002; Holden and Mace 2003). In addition, archaeologists who study cultural evolution have a similar modus operandi as paleobiologists who study genetic evolution (Mesoudi, Whiten, and Laland 2006). Both attempt to reconstruct lineages of artifacts and their carriers. Like paleobiology, archaeology assumes that when analogy can be ruled out, similarity implies causal connection by inheritance (O'Brian and Lyman 2000). Like biogeography's study of the spatial distribution of organisms (Brown

and Lomolino 1998), behavioral ecology studies the interaction of ecological, historical, and geographical factors that determine the distribution of cultural forms across space and time (Smith and Winterhalder 1992).

Perhaps the most common critique of the analogy between genetic and cultural evolution is that the gene is a well-defined, discrete, independently reproducing and mutating entity, whereas the boundaries of the unit of culture are ill-defined and overlapping. In fact, however, this view of the gene is simply outdated. Overlapping, nested, and movable genes discovered in the past 35 years have some of the fluidity of cultural units, whereas quite often the boundaries of a cultural unit (a belief, icon, word, technique, stylistic convention) are quite delimited and specific. Similarly, alternative splicing, nuclear and messenger RNA editing, cellular protein modification, and genomic imprinting, which are quite common, quite undermine the standard view of the insular gene producing a single protein and support the notion of genes having variable boundaries and having strongly context-dependent effects.

Dawkins added a second fundamental mechanism of epigenetic information transmission in *The Extended Phenotype* (1982), noting that organisms can directly transmit environmental artifacts to the next generation in the form of such constructs as beaver dams, beehives, and even social structures (e.g., mating and hunting practices). The phenomenon of a species creating an important aspect of its environment and stably transmitting this environment across generations, known as *niche construction*, it a widespread form of epigenetic transmission (Odling-Smee, Laland, and Feldman 2003). Moreover, niche construction gives rise to what might be called a *gene-environment coevolutionary process* since a genetically induced environmental regularity becomes the basis for genetic selection, and genetic mutations that give rise to mutant niches will survive if they are fitness-enhancing for their constructors.

An excellent example of gene-environment coevolution is seen in the honey bee, for which the origin of its eusociality likely lay in the high degree of relatedness fostered by haplodiploidy but still which persists in modern species despite the fact that relatedness in the hive is generally quite low, because of multiple queen matings, multiple queens, queen deaths, and the like (Gadagkar 1991; Seeley 1997; Wilson and Holldobler 2005).[2] The

[2] A *social* species is one that has a division of labor and cooperative behavior. A *eusocial* species is a social species that has a reproductive division of labor; i.e., some females, such as queen bees, produce offspring, while other females, such as worker bees, raise

social structure of the hive is transmitted epigenetically across generations, and the honey bee genome is an adaptation to the social structure laid down in the distant past.

Gene-culture coevolution in humans is a special case of gene-environment coevolution in which the environment is culturally constituted and transmitted (Feldman and Zhivotovsky 1992). The key to the success of our species in the framework of the hunter-gatherer social structure in which we evolved is the capacity of unrelated, or only loosely related, individuals to cooperate in relatively large egalitarian groups in hunting and territorial acquisition and defense (Boehm 2000; Richerson and Boyd 2004). While contemporary biological and economic theory have attempted to show that such cooperation can be effected by self-regarding rational agents (Trivers 1971; Alexander 1987; Fudenberg, Levine, and Maskin 1994), the conditions under which this is the case are highly implausible even for small groups (Boyd and Richerson 1988; Gintis 2005). Rather, the social environment of early humans was conducive to the development of prosocial traits, such as empathy, shame, pride, embarrassment, and reciprocity, without which social cooperation would be impossible.

Neuroscientific studies exhibit clearly the genetic basis for moral behavior. Brain regions involved in moral judgments and behavior include the prefrontal cortex, the orbitalfrontal cortex, and the superior temporal sulcus (Moll et. al 2005). These brain structures are virtually unique to, or most highly developed in humans and are doubtless evolutionary adaptations (Schulkin 2000). The evolution of the human prefrontal cortex is closely tied to the emergence of human morality (Allman, Hakeem, and Watson 2002). Patients with focal damage to one or more of these areas exhibit a variety of antisocial behaviors, including the absence of embarrassment, pride, and regret (Beer et. al al 2003; Camille 2004), and sociopathic behavior (Miller et. al al 1997). There is a likely genetic predisposition underlying sociopathy. Sociopaths comprise 3% to 4% of the male population, but they account for between 33% and 80% of the population of chronic criminal offenders in the United States (Mednick et. al al 1977).

It is clear from this body of empirical information that culture is directly encoded in the human brain, which of course is the central claim of gene-culture coevolutionary theory.

the queen's offspring. A haplodiploid species is one in which one sex inherits from both parents, while the other sex inherits only from one parent.

12.2 Biological and Cultural Dynamics

The analysis of living systems includes one concept that is not analytically represented in the natural sciences: that of a *strategic interaction* in which the behavior of agents is derived by assuming that each is choosing a *best response* to the actions of other agents. The study of systems in which agents choose best responses and in which such responses evolve dynamically is called *evolutionary game theory*.

A *replicator* is a physical system capable of drawing energy and chemical building blocks from its environment to make copies of itself. Chemical crystals, such as salt, have this property, but biological replicators have the additional ability to assume a myriad of physical forms based on the highly variable sequencing of their chemical building blocks. Biology studies the dynamics of such complex replicators using the evolutionary concepts of replication, variation, mutation, and selection (Lewontin 1974).

Biology plays a role in the behavioral sciences much like that of physics in the natural sciences. Just as physics studies the elementary processes that underlie all natural systems, so biology studies the general characteristics of survivors of the process of natural selection. In particular, genetic replicators, the epigenetic environments to which they give rise, and the effect of these environments on gene frequencies account for the characteristics of species, including the development of individual traits and the nature of intraspecific interaction. This does not mean, of course, that behavioral science in any sense can be *reduced* to biological laws. Just as one cannot deduce the character of natural systems (e.g., the principles of inorganic and organic chemistry, the structure and history of the universe, robotics, plate tectonics) from the basic laws of physics, similarly, one cannot deduce the structure and dynamics of complex life forms from basic biological principles. But, just as physical principles inform model creation in the natural sciences, so must biological principles inform all the behavioral sciences.

Within population biology, evolutionary game theory has become a fundamental tool. Indeed, evolutionary game theory is basically population biology with frequency-dependent fitnesses. Throughout much of the twentieth century, classical population biology did not employ a game-theoretic framework (Fisher 1930; Haldane 1932; Wright 1931). However, Moran (1964) showed that Fisher's fundamental theorem, which states that as long as there is positive genetic variance in a population, fitness increases over time, is false when more than one genetic locus is involved. Eshel and Feldman (1984) identified the problem with the population genetics model

in its abstraction from mutation. But how do we attach a fitness value to a mutant? Eshel and Feldman (1984) suggested that payoffs be modeled game-theoretically on the phenotypic level and that a mutant gene be associated with a strategy in the resulting game. With this assumption, they showed that under some restrictive conditions, Fisher's fundamental theorem could be restored. Their results were generalized by Liberman (1988), Hammerstein and Selten (1994), Hammerstein (1996), Eshel, Feldman, and Bergman (1998), and others.

The most natural setting for genetic and cultural dynamics is game-theoretic. Replicators (genetic and/or cultural) endow copies of themselves with a repertoire of strategic responses to environmental conditions, including information concerning the conditions under which each is to be deployed in response to the character and density of competing replicators. Genetic replicators have been well understood since the rediscovery of Mendel's laws in the early twentieth century. Cultural transmission also apparently occurs at the neuronal level in the brain, in part through the action of *mirror neurons* (Williams et. al al 2001; Rizzolatti et. al al 2002; Meltzhoff and Decety 2003). Mutations include replacement of strategies by modified strategies, and the "survival of the fittest" dynamic (formally called a *replicator dynamic*) ensures that replicators with more successful strategies replace those with less successful strategies (Taylor and Jonker 1978).

Cultural dynamics, however, do not reduce to replicator dynamics. For one thing, the process of switching from lower- to higher-payoff cultural norms is subject to error, and with some positive frequency, lower-payoff forms can displace higher-payoff forms (Edgerton 1992). Moreover, cultural evolution can involve a conformist predisposition (Henrich and Boyd 1998; Henrich and Boyd 2001; Guzman, Sickert, and Rowthorn 2007), as well as oblique and horizontal transmission (Cavalli-Sforza and Feldman 1981; Gintis 2003b).

12.3 The Theory of Norms: The Sociological Model

Complex social systems generally have a division of labor, with distinct social positions occupied by individuals specially prepared for their roles. For instance, a beehive has workers, drones, and queens, and workers can be nurses, foragers, or scouts. Preparation for roles is by gender and larval nutrition. Modern human society has a division of labor characterized by

dozens of specialized *roles*, appropriate behavior within which is given by *social norms*, and individuals are *actors* who are motivated to fulfill these roles through a combination of *material incentives* and *normative commitments*.

The centrality of culture in the social division of labor was clearly expressed by Emile Durkheim (1933 [1902]), who stressed that the great multiplicity of roles (which he called *organic solidarity*) required a commonality of beliefs (which he called *collective consciousness*) that would permit the smooth coordination of actions by distinct individuals. This theme was developed by Talcott Parsons (1937), who used his knowledge of economics to articulate a sophisticated model of the interaction between the situation (role) and its inhabitant (actor). The actor/role approach to social norms was filled out by Erving Goffman (1959), among others.

The social role has both normative and positive aspects. On the positive side, the payoffs—rewards and penalties—associated with a social role must provide the appropriate incentives for actors to carry out the duties associated with the role. This requirement is most easily satisfied when these payoffs are independent of the behavior of agents occupying other roles. However, this is rarely the case. In general, as developed in chapter 7, social roles are deeply interdependent and can be modeled as the strategy sets of players in an epistemic game, the payoffs to which are precisely these rewards and penalties, the choices of actors then forming a correlated equilibrium for which the required commonality of beliefs is provided by a society's common culture. This argument provides an analytical link uniting the actor/role framework in sociological theory with game-theoretic models of cooperation in economic theory.

Appropriate behavior in a social role is given by a *social norm* that specifies the duties, privileges, and normal behavior associated with the role. In the first instance, social norms have an instrumental character devoid of normative content, serving merely as informational devices that coordinate the behavior of rational agents (Lewis 1969; Gauthier 1986; Binmore 2005; Bicchieri 2006). However, in most cases, high level performance in a social role requires that the actor have a *personal commitment* to role performance that cannot be captured by the self-regarding "public" payoffs associated with the role (see chapter 7 and Conte and Castelfranchi, 1999). . This is because (a) actors may have private payoffs that conflict with the role's public payoffs, inducing them to behave counter to proper role-performance (e.g., corruption, favoritism, aversion to specific tasks); (b)

the signal used to determine the public payoffs may be inaccurate and unreliable (e.g., the performance of a teacher, physician, scientist, or business executive cannot be fully objectively assessed at reasonable cost); and (c) the public payoffs required to gain compliance by self-regarding actors may be higher than those required when there is at least partial reliance upon the personal commitment of role incumbents; i.e., it may be less costly to use personally committed rather than purely materially motivated agents when performance cannot be easily measured (Bowles 2008). In such cases, self-regarding actors who treat social norms purely instrumentally behave in a socially inefficient manner (§6.3, 6.4).

The normative aspect of social roles flows from these considerations. First, to the extent that social roles are considered legitimate by incumbents, they place an intrinsic ethical value on role performance. We call this the *normative predisposition* associated with role occupancy (see chapter 7). Second, human ethical predispositions include *character virtues*, such as honesty, trustworthiness, promise keeping, and obedience, that may increase the value of conforming to the duties associated with role incumbency (§3.12). Third, humans are also predisposed to care about the esteem of others even when there can be no future reputational repercussions (Masclet et. al al 2003) and take pleasure in punishing others who have violated social norms (Fehr and Fischbacher 2004). These ethical traits by no means contradict rationality (§12.5), because individuals trade off these values against material reward, and against each other, just as described in the economic theory of the rational actor (Andreoni and Miller 2002; Gneezy and Rustichini 2000).

The sociopsychological theory of norms can thus resolve the contradictions between the sociological and economic models of social cooperation, retaining the analytical clarity of game theory and the rational actor model while incorporating the collective, normative, and cultural characteristics stressed in psychosocial models of norm compliance.

12.4 Socialization and the Internalization of Norms

Society is held together by *moral values* that are transmitted from generation to generation by the process of *socialization*. These values are instantiated through the *internalization of norms* (Parsons 1967; Grusec and Kuczynski 1997; Nisbett and Cohen 1996; Rozin et. al 1999), a process in which the initiated instill values into the uninitiated (usually the younger

generation) through an extended series of personal interactions, relying on a complex interplay of affect and authority. Through the internalization of norms, initiates are supplied with moral values that induce them to conform to the duties and obligations of the role positions they expect to occupy. The internalization of norms, of course, presupposes a genetic predisposition to moral cognition that can be explained only by gene-culture coevolution.

Internalized norms are accepted not as instruments for achieving other ends but rather as *arguments in the preference function that the individual maximizes*. For instance, an individual who has internalized the value of speaking truthfully does so even in cases where the net payoff to speaking truthfully would otherwise be negative. Such fundamental human emotions as shame, guilt, pride, and empathy are deployed by the well-socialized individual to reinforce these prosocial values when tempted by the immediate pleasures of such deadly sins as anger, avarice, gluttony, and lust. It is tempting to treat some norms as constraints rather than objectives, but virtually all norms are violated by individuals under some conditions, indicating that there are tradeoffs, such as those explored in §3.12 and §3.4, that could not exist were norms merely constraints on action.

The human openness to socialization is perhaps the most powerful form of epigenetic transmission found in nature. This epigenetic flexibility in considerable part accounts for the stunning success of the species *H. sapiens* because when individuals internalize a norm, the frequency of the desired behavior is higher than if people follow the norm only instrumentally—i.e., when they perceive it to be in their best interest to do so on other grounds. The increased incidence of prosocial behaviors is precisely what permits humans to cooperate effectively in groups (Gintis et. al 2005).

There are, of course, limits to socialization (Tooby and Cosmides 1992; Pinker 2002), and it is imperative to understand the dynamics of the emergence and abandonment of particular values, which in fact depend on their contribution to fitness and well-being, as economic and biological theory would suggest (Gintis 20031,b). Moreover, there are often swift, society-wide value changes that cannot be accounted for by socialization theory (Wrong 1961; Gintis 1975). However, socialization theory has an important place in the general theory of culture, strategic learning, and moral development.

One of the more stunning indications of the disarray of the behavioral sciences is the fact that the internalization of norms does not appear in the economic and biological models of human behavior.

12.5 Rational Choice: The Economic Model

General evolutionary principles suggest that individual decision making for members of a species can be modeled as optimizing a preference function. Natural selection leads the content of preferences to reflect biological fitness. The principle of expected utility extends this optimization to stochastic outcomes. The resulting model is called the *rational actor model* in economics, although there is some value to referring to it as the *beliefs, preferences, and constraints* (BPC) model, thus avoiding the often misleading connotations attached to the term "rational."

For every constellation of sensory inputs, each decision taken by an organism generates a probability distribution over outcomes, the expected value of which is the *fitness* associated with that decision. Since fitness is a scalar variable, for each constellation of sensory inputs, each possible action the organism might take has a specific fitness value, and organisms whose decision mechanisms are optimized for this environment choose the available action that maximizes this value. This argument was presented verbally by Darwin (1872) and is implicit in the standard notion of "survival of the fittest," but formal proof is recent (Grafen 1999, 2000, 2002). The case with frequency-dependent (nonadditive genetic) fitness has yet to be formally demonstrated, but the informal arguments are compelling.

Given the state of its sensory inputs, if an organism with an optimized brain chooses action A over action B when both are available, and chooses action B over action C when both are available, then it will also choose action A over action C when both are available. Thus, choice consistency follows from basic evolutionary dynamics. The rational actor model is often presented as though it applies only when actors possess extremely powerful information-processing capacities. As we saw in chapter 1, in fact, the basic model depends only on choice consistency, the expected utility theorem being considerably more demanding.

Four *caveats* are in order. First, individuals do not consciously maximize something called "utility," or anything else. Second, individual choices, even if they are self-regarding (e.g., personal consumption) are not necessarily welfare-enhancing. Third, preferences must have some stability across time to be theoretically useful, but preferences are ineluctably a function of an individual's *current state*, and beliefs can change dramatically in response to immediate sensory and social experiences. Finally, beliefs need not be correct nor need they be updated correctly in the face of

new evidence, although Bayesian assumptions concerning updating can be made part of consistency in elegant and compelling ways (Jaynes 2003).

The rational actor model is the cornerstone of contemporary economic theory and in the past few decades has become the heart of the biological modeling of animal behavior (Real 1991; Alcock 1993; Real and Caraco 1986). Economic and biological theory thus have a natural affinity: the choice consistency on which the rational actor model of economic theory depends is rendered plausible by evolutionary theory, and the optimization techniques pioneered in economics are routinely applied and extended by biologists in modeling the behavior of nonhuman organisms. I suggest below that this is due to the *routine* choice paradigm that applies in economics and biology, as opposed to the *deliberative* choice paradigm that applies in cognitive psychology.

Perhaps the most pervasive critique of the BPC model is that put forward by Herbert Simon (1982), holding that because information processing is costly and humans have a finite information-processing capacity, individuals *satisfice* rather than *maximize* and hence are only *boundedly rational*. There is much substance to this view, including the importance of including information-processing costs and limited information in modeling choice behavior and recognizing that the decision on how much information to collect depends on unanalyzed subjective priors at some level (Winter 1971; Heiner 1983). Indeed, from basic information theory and quantum mechanics, it follows that *all rationality is bounded*. However, the popular message taken from Simon's work is that we should reject the BPC model. For instance, the mathematical psychologist D. H. Krantz (1991) asserts, "The normative assumption that individuals *should* maximize *some* quantity may be wrong. ...People do and should act as *problem solvers*, not *maximizers*." This is incorrect. In fact, as long as individuals are involved in routine choice (see §12.13) and hence have consistent preferences, they can be modeled as maximizing an objective function subject to constraints.

This point is lost on even such capable researchers as Gigerenzer and Selten (2001), who reject the "optimization subject to constraints" method on the grounds that individuals do not in fact solve optimization problems. However, just as billiards players do not solve differential equations in choosing their shots, so decision makers do not solve Lagrangian equations, even though in both cases we may use such optimization models to describe their behavior. Of course, as stressed by Gigerenzer and Selten (2001), from an analytical standpoint, generalizing the rational actor model

may not be the best way to capture the heuristics of decision making in particular areas.

12.6 Deliberative Choice: The Psychological Model

The psychological literature on decision making is rich and multifaceted, traditional approaches being augmented in recent years by neural net theory and evidence from neuroscientific data on brain functioning (Kahneman, Slovic, and Tversky 1982; Baron 2007; Oaksford and Chater 2007; Hinton and Sejnowski 1999; Newell, Lagnado, and Shanks 2007; Juslin and Montgomery 1999; Bush and Mosteller 1955; Gigerenzer and Todd 1999; Betch and Haberstroh 2005; Koehler and Harvey 2004). There does not yet exist a unitary model underlying the psychological understanding of judgment and decision making, doubtless because the mental processes involved are so varied and complex.

The sorts of decision making studied by psychologists include the formation of long-term goals, which are evaluated according to the value if attained, the range of probable costs, and the probability of goal attainment. All three dimensions of goal formation have inherent uncertainties, so among the strategies of goal choice is the formation of subgoals with the aim of reducing these uncertainties. The most complex of human decisions tend to involve goals that arise infrequently in the course of a life, such as choosing a career, whether to marry and to whom, how many children to have, and how to deal with a health threat, where the scope for learning from mistakes is narrow. Psychologists also study how people make decisions based on noisy single- or multidimensional data under conditions of trial-and-error learning.

The difficulty in modeling a deliberative choice is exacerbated by the fact that, because of the complexity of such decisions, much human decision making has a distinctly group dynamic, in which some individuals experiment and other imitate the more successful of the experimenters (Bandura 1977). This dynamic cannot be successfully modeled on the individual level.

By contrast, the rational actor model applies to choice situations where ambiguities are absent, the choice set is clearly delineated, and the payoffs are unmediated, so that no deliberation is involved beyond the comparison of feasible alternatives. Accordingly, most psychologists working in this area accept the rational actor model as the appropriate model of choice be-

havior in this realm of routine choice, yet recognize that there is no obvious way to extend the model to the more complex situations they study. For instance, Newell, Lagnado, and Shanks (2007) assert, "We view judgment and decision making as often exquisitely subtle and well-tuned to the world, especially in situations where we have the opportunity to respond repeatedly under similar conditions where we can learn from feedback." (p. 2)

There is thus no deep conceptual divide between the psychological approach to decision making and the economic approach. While in some important areas, human decision makers appear to violate the consistency condition for rational choice, in virtually all such cases, as we suggested in §1.9, consistency can be restored by assuming that the current state of the agent is an argument of the preference structure. Another possible challenge to preference consistency is preference reversal in the choice of lotteries. Lichtenstein and Slovic (1971) were the first to find that in many cases, individuals who prefer lottery A to lottery B are nevertheless willing to take less money for A than for B. Reporting this to economists several years later, Grether and Plott (1979) asserted, "A body of data and theory has been developed...[that] are simply inconsistent with preference theory...(p. 623). These preference reversals were explained several years later by Tversky, Slovic, and Kahneman (1990) as a bias toward the higher probability of winning in a lottery choice and toward the higher maximum amount of winnings in monetary valuation. However, the phenomenon has been documented only when the lottery pairs A and B are so close in expected value that one needs a calculator (or a quick mind) to determine which would be preferred by an expected value maximizer. For instance, in Grether and Plott (1979) the average difference between expected values of comparison pairs was 2.51% (calculated from table 2, p. 629). The corresponding figure for Tversky, Slovic, and Kahneman (1990) was 13.01%. When the choices are so close to indifference, it is not surprising that inappropriate cues are relied upon to determine choice, as would be suggested by the heuristics and biases model (Kahneman, Slovic, and Tversky 1982) favored by behavioral economists and psychologists.

The expected utility model (§1.5) is closer to the concerns of psychologists because it deals with uncertainty in a fundamental way, and applying Bayes' rule certainly may involve complex deliberations. The Ellsberg paradox is an especially clear example of the failure of the probability reasoning behind the expected utility model. Nevertheless the model has a considerable body of empirical support, so the basic modeling issue is to be

able to say clearly when the expected utility theorem is likely to be violated, and to supply an alternative model outside this range (Newell, Lagnado, and Shanks 2007; Oaksford and Chater 2007).

I conclude that there should be a basic synergy between the rational actor model when dealing with routine choice and the sorts of models developed by psychologists to explain complex human deliberation, goal formation, and learning.

12.7 Application: Addictive Behavior

Substance abuse appears to be irrational. Abusers are time inconsistent and their behavior is welfare-reducing. Moreover, even draconian increases in the penalties for illicit substance use lead to the swelling of prison populations rather than abandonment of the sanctioned activity. Because rational actors generally trade off among desired goals, this curious phenomenon has led some researchers to reject the BPC model out of hand.

However, the BPC model remains the most potent tool for analyzing substance abuse on a societywide level. The most salient target of the critics has been the "rational addiction" model of Becker and Murphy (1988). While this model does have some shortcomings, its use of the rational actor model is not among them. Indeed, empirical research supports the contention that illicit drugs respond normally to market forces. For instance, Saffer and Chaloupka (1999) estimated the price elasticities of heroin and cocaine using a sample of 49,802 individuals from the National Household Survey of Drug Abuse to be 1.70 and 0.96, respectively. These elasticities are in fact quite high. Using these estimates, the authors judge that lower prices flowing from the legalization of these drugs would lead to an increase of about 100% and 50% in the quantities of heroin and cocaine consumed, respectively.

How does this square with the observation that draconian punishments do not squelch the demand altogether? Gruber and Köszegi (2001), who use the rational actor model but do not assume time consistency, explain this by showing that drug users exhibit the commitment and self-control problems typical of time-inconsistent agents, for whom the possible future penalties have a highly attenuated deterrent value in the present. This behavior may be welfare-reducing, but the rational actor model does not presume that preferred outcomes are necessarily welfare-improving.

12.8 Game Theory: The Universal Lexicon of Life

Game theory is a logical extension of evolutionary theory. To see this, suppose there is only one replicator, deriving its nutrients and energy from nonliving sources. The replicator population will then grow at a geometric rate until it presses upon its environmental inputs. At that point, mutants that exploit the environment more efficiently will outcompete their less efficient conspecifics and with input scarcity, mutants will emerge that "steal" from conspecifics who have amassed valuable resources. With the rapid growth of such predators, mutant prey will devise means of avoiding predation, and predators will counter with their own novel predatory capacities. In this manner, strategic interaction is born from elemental evolutionary forces. It is only a conceptual short step from this point to cooperation and competition among cells in a multicellular body, among conspecifics who cooperate in social production, between males and females in a sexual species, between parents and offspring, and among groups competing for territorial control.

Historically, game theory did not emerge from biological considerations but rather from strategic concerns in World War II (Von Neumann and Morgenstern 1944; Poundstone 1992). This led to the widespread caricature of game theory as applicable only to static confrontations of rational self-regarding agents possessed of formidable reasoning and information-processing capacity. Developments within game theory in recent years, however, render this caricature inaccurate.

Game theory has become the basic framework for modeling animal behavior (Maynard Smith 1982; Alcock 1993; Krebs and Davies 1997) and thus has shed its static and hyperrationalistic character, in the form of evolutionary game theory (Gintis 2009). Evolutionary and behavioral game theory do not require the formidable information-processing capacities of classical game theory, so disciplines that recognize that cognition is scarce and costly can make use of game-theoretic models (Young 1998; Gintis 2009; Gigerenzer and Selten 2001). Thus, agents may consider only a restricted subset of strategies (Winter 1971; Simon 1972), and they may use rule-of-thumb heuristics rather than maximization techniques (Gigerenzer and Selten 2001). Game theory is thus a generalized schema that permits the precise framing of meaningful empirical assertions but imposes no particular structure on the predicted behavior.

12.9 Epistemic Game Theory and Social Norms

Economics and sociology have highly contrasting models of human interaction. Economics traditionally considers individuals to be rational, self-regarding payoff maximizers, while sociology considers individuals to be highly socialized, other-regarding, moral agents who strive to fill social roles and whose self-esteem depends on the approbation of others. The project of unifying the behavioral sciences must include a resolution of these inconsistencies in a manner that preserves the key insights of each.

Behavioral game theory helps us adjudicate these disciplinary differences, providing experimental data supporting the sociological stress on moral values and other-regarding preferences, and also supports the economic stress on rational payoff maximization. For instance, most individuals care about reciprocity and fairness as well as personal gain (Gintis et. al 2005), value such character virtues as honesty for their own sake (Gneezy 2005), care about the esteem of others even when there can be no future reputational repercussions (Masclet et. al 2003), and take pleasure in punishing others who have hurt them (deQuervain et. al 2004). Moreover, as suggested by socialization theory, individuals have consistent values, based on their particular sociocultural situations, that they apply in the laboratory even in one-shot games under conditions of anonymity (Henrich et. al 2004; Henrich et. al 2006). This body of evidence suggests that sociologists would benefit from reincorporating the rational actor model into sociological theory, and economists broaden their concept of human preferences.

A second discrepancy between economics and sociology concerns the contrasting claims of game theory and the sociopsychological theory of norms in explaining social cooperation. Our exposition of this area in chapter 7 can be interpreted in the larger context of the unity of the behavioral sciences as follows.

The basic model of the division of labor in economic theory is the Walrasian general equilibrium model, according to which a system of flexible prices induces firms and individuals to supply and demand goods and services in such amounts that all markets clear in equilibrium (Arrow and Debreu 1954). However, this model assumes that all contracts among individuals can be costlessly enforced by a third party, such as the judicial system. In fact, however, many critical forms of social cooperation are not mediated by a third-party enforceable contract but rather take the form of repeated interactions in which an informal, but very real, threat of rupturing the relationship is used to induce mutual cooperation (Fudenberg, Levine,

and Maskin 1994; Ely and Välimäki 2002). For instance, an employer hires a worker who works hard under the threat of dismissal not the threat of an employer lawsuit.

Repeated game theory thus steps in for economists to explain forms of face-to-face cooperation that do not reduce to simple price-mediated market exchanges. Repeated game theory shows that in many cases the activity of many individuals can be coordinated, in the sense that there is a Nash equilibrium ensuring that no self-regarding player can gain by deviating from the strategy assigned to him by the equilibrium, assuming other players also use the strategies assigned to them (§10.4). If this theory were adequate, which most economists believe is the case, then there would be no role for the sociopsychological theory of norms, and sociological theory would be no more that a thick description of a social mechanism analytically accounted for by repeated game theory.

However, repeated game theory with self-regarding agents does not solve the problem of social cooperation (§10.6). When the group consists of more than two individuals and the signal indicating how well a player is performing his part is imperfect and private (i.e., players receive imperfectly correlated signals about another player's behavior), the efficiency of cooperation may be quite low, and the roles assigned to each player will be extremely complex mixed strategies that players have no incentive to use (§10.5). As we suggested in chapter 7, the sociopsychology of norms can step in at this point to provide mechanisms that induce individuals to play their assigned parts. A social norm may provide the rules for each individual in the division of labor, players may have a general predilection for honesty that allows them to consolidate their private signals concerning another player's behavior into a public signal that can be the bases for coordinated collective punishment and reward, and players may have a personal normative predisposition towards following the social roles assigned to them. The sociological and economic forces thus complement rather than contradict one another.

A central analytical contribution to this harmonization of economics and sociology was provided by Robert Aumann (1987), who showed that the natural concept of equilibrium in game theory for rational actors who share common beliefs is not the Nash equilibrium but the correlated equilibrium. A correlated equilibrium is the Nash equilibrium in the game formed by adding to the original game a new player, whom I call the *choreographer* (Aumann calls this simply a "correlating device"), who samples the proba-

bility distribution given by the players (common) beliefs and then instructs each player what action to take. The actions recommended by the choreographer are all best responses to one another, conditional on their having been simultaneously ordered by the choreographer, so self-regarding players can do no better than to follow the choreographer's advice.

Sociology, and more generally sociobiology (see chapter 11), then come in not only by supplying the choreographer, in the form of a complex of social norms, but also by supplying cultural theory to explain why players might have a common set of beliefs, without which the correlated equilibrium would not exist. Cognitive psychology explains the normative predisposition that induces players to take the advice of the choreographer (i.e., to follow the social norm) when in fact there might be many other actions with equal, or even higher, payoff that the player might have an inclination to choose.

12.10 Society as a Complex Adaptive System

The behavioral sciences advance not only by developing analytical and quantitative models but also by accumulating historical, descriptive, and ethnographic evidence that pays heed to the detailed complexities of life in the sweeping array of wondrous forms that nature reveals to us. Historical contingency is a primary focus for many students of sociology, anthropology, ecology, biology, politics, and even economics. By contrast, the natural sciences have found little use for narrative alongside analytical modeling.

The reason for this contrast between the natural and the behavioral sciences is that *living systems are generally complex, dynamic adaptive systems* with emergent properties that cannot be fully captured in analytical models that attend only to local interactions. The hypothetico-deductive methods of game theory, the BPC model, and even gene-culture coevolutionary theory must therefore be complemented by the work of behavioral scientists who adhere to more historical and interpretive traditions, as well as that of researchers who use agent-based programming techniques to explore the dynamic behavior of approximations to real-world complex adaptive systems.

A *complex system* consists of a large population of similar entities (in our case, human individuals) who interact through regularized channels (e.g., networks, markets, social institutions) with significant stochastic elements, without a system of centralized organization and control (i.e., if there is a

state, it controls only a fraction of all social interactions and is itself a complex system). A complex system is *adaptive* if it undergoes an evolutionary (genetic, cultural, agent-based, or other) process of reproduction, mutation, and selection (Holland 1975). To characterize a system as complex adaptive does not explain its operation and does not solve any problems. However, it suggests that certain modeling tools are likely to be effective that have little use in a noncomplex system. In particular, the traditional mathematical methods of physics and chemistry must be supplemented by other modeling tools such as agent-based simulation and network theory.

The stunning success of modern physics and chemistry lies in their ability to avoid or control emergence. The experimental method in natural science is to create highly simplified laboratory conditions under which modeling becomes analytically tractable. Physics is no more effective than economics or biology in analyzing complex real-world phenomena in situ. The various branches of engineering (electrical, chemical, mechanical) are effective because they re-create in everyday life artificially controlled, noncomplex, nonadaptive environments in which the discoveries of physics and chemistry can be directly applied. This option is generally not open to most behavioral scientists, who rarely have the opportunity of "engineering" social institutions and cultures.

12.11 Counterpoint: Biology

Biologists are generally comfortable with three of the five principles laid out in the introduction to this chapter. Only gene-culture coevolution and the sociopsychology of norms have generated significant opposition.

Gene-culture coevolutionary theory has been around only since the 1980s and applies to only one species—*H. sapiens*. Not surprisingly, many sociobiologists have been slow to adopt it and have deployed a formidable array of population biology concepts toward explaining human sociality in more familiar terms—especially kin selection (Hamilton 1964) and reciprocal altruism (Trivers 1971). The explanatory power of these models convinced a generation of researchers that what appears to be altruism— personal sacrifice on the behalf of others—is really just long-run self-interest, and that elaborate theories drawn from anthropology, sociology, and economics are unnecessary to explain human cooperation and conflict.

Richard Dawkins, for instance, in *The Selfish Gene* (1989 [1976]), asserts, "We are survival machines—robot vehicles blindly programmed to preserve

the selfish molecules known as genes.... This gene selfishness will usually give rise to selfishness in individual behavior." Similarly, in *The Biology of Moral Systems* (1987), R. D. Alexander asserts, "Ethics, morality, human conduct, and the human psyche are to be understood only if societies are seen as collections of individuals seeking their own self-interest...." (p. 3) In a similar vein, Michael Ghiselin (1974) writes, "No hint of genuine charity ameliorates our vision of society, once sentimentalism has been laid aside. What passes for cooperation turns out to be a mixture of opportunism and exploitation.... Scratch an altruist, and watch a hypocrite bleed" (p. 247)

Evolutionary psychology, which has been a major contributor to human sociobiology, has incorporated the kin selection/reciprocal altruism perspective into a broadside critique of the role of culture in society (Barkow, Cosmides, and Tooby 1992) and of the forms of group dynamics upon which gene-culture coevolution depends (Price, Cosmides, and Tooby 2002). I believe these claims have been effectively refuted (Richerson and Boyd 2004; Gintis et. al al 2009), although the highly interesting debate in population biology concerning group selection has been clarified but not completely resolved (Lehmann and Keller 2006; Lehmann et. al al 2007; Wilson and Wilson 2007).

12.12 Counterpoint: Economics

Economists generally believe in methodological individualism, a doctrine claiming that all social behavior can be explained by strategic interactions among agents. Were this correct, gene-culture coevolution would be unnecessary, complexity theory would be irrelevant, and the sociopsychological theory of norms could be derived from game theory. We concluded in chapter 8, however, that methodological individualism is contradicted by the evidence.

Economists also generally reject the idea of society as a complex adaptive system, on grounds that we may yet be able to tweak the Walrasian general equilibrium framework, suitably fortified by sophisticated mathematical methods, so as to explain macroeconomic activity. In fact, there has been virtually no progress in general equilibrium theory since the mid-twentieth-century existence proofs (Arrow and Debreu 1954). Particularly noteworthy has been the absence of any credible stability model (Fisher 1983). Indeed, the standard models predict price instability and chaos (Saari

1985; Bala and Majumdar 1992). Moreover, analysis of excess demand functions suggests that restrictions on preferences are unlikely to entail the stability of Walrasian price dynamics (Sonnenschein 1972, 1973; Debreu 1974; Kirman and Koch 1986).

My response to this sad state of affairs has been to show that agent-based models of generalized exchange, based on the notion that the economy is a complex nonlinear dynamical system, exhibit a high degree of stability and efficiency (Gintis 2006, 2007a). There does not appear to be any serious doctrinal impediment to the use of agent-based modeling in economics.

12.13 Counterpoint: Psychology

Decision theory, based on the rational actor model, represents one of the great scientific achievements of all time, beginning with Bernoulli and Pascal in the seventeenth and eighteenth centuries and culminating in the work of Ramsey, de Finetti, Savage, and von Neumann and Morgenstern in the early and middle years of the twentieth century. Its preeminence in the behavioral disciplines that deal with human choice, especially its position as the keystone of modern economic theory, however, has led to an extreme level of empirical scrutiny of decision theory. Because I include the rational actor model as one of my five organizing principles for the unification of the behavioral sciences, this critique deserves careful consideration.

The most salient critique has taken inspiration from the brilliant series of experiments by Daniel Kahneman and Amos Tversky. These researchers have documented several key and systematic divergences between the normative principle of decision theory and the actual choices of intelligent, educated individuals (see chapter 1). Such phenomena as loss aversion, the base rate fallacy, framing effects, and the conjunction fallacy must be added to the traditional paradoxes of Allais and Ellsberg as representing fundamental aspects of human decision making that fall outside the purview of traditional decision theory (§1.7).

Psychologists have used these contributions improperly to mount a sustained attack on the rational actor model, leading many researchers to reject traditional decision theory and seek alternatives lying quite outside the rational actor tradition, in such areas as computer modeling of neural nets and neuroscientific studies of brain functioning. This dismissal of traditional decision theory may be emotionally satisfying, but it is immature, short-sighted, and scientifically destructive. There is no alternative to the

traditional decision-theoretic model on the horizon, and there is not likely to be one, for one simple reason: the theory is mostly correct, and where it fails, the principles accounting for failure are complementary to, rather than destructive of, the standard theory. For instance, the documented inconsistencies in the traditional rational actor model can be handled effectively by assuming that the preference function has the current state of the individual as an argument, so all assessments are of deviations from the status quo ante. Prospect theory, for which Kahneman was awarded the Nobel prize, is precisely of this form, as is the treatment of time inconsistency and regret phenomena. In other cases, by assuming that individuals have other-regarding preferences (which laboratory evidence strongly supports), we rupture the traditional prejudice that rationality implies selfishness.

My suggestion for resolving the conflict between psychological and economic models of decision making has four points. First, the two disciplines should recognize the distinction between deliberative and routine decision making. Second, psychology should introduce the evolution of routine decision making into its core framework, based on the principle that the brain is a fitness-enhancing adaptation. Third, deliberative decision making is an adaptation to the increased social complexity of primates and hominid groups. Finally, routine decision making shades into deliberative decision making under conditions that are only imperfectly known but are of great potential importance for understanding human choice.

12.14 The Behavioral Disciplines Can Be Unified

In this chapter, I have proposed five analytical tools that together serve to provide a common basis for the behavioral sciences. These are gene-culture coevolution, the sociopsychological theory of norms, game theory, the rational actor model, and complexity theory. While there are doubtless formidable scientific issues involved in providing the precise articulations between these tools and the major conceptual tools of the various disciplines, as exhibited, for instance, in harmonizing the socio-psychological theory of norms and repeated game theory, these intellectual issues are likely to be dwarfed by the sociological issues surrounding the semifeudal nature of modern behavioral disciplines, which renders even the most pressing reform a monumental enterprise. If these institutional obstacles can be overcome, the behavioral disciplines can be unified.