A Framework for the Unification of the Behavioral Sciences

Herbert Gintis

May 6, 2006

Abstract

The various behavioral disciplines model human behavior in distinct and incompatible ways. Yet, recent theoretical and empirical developments have created the conditions for rendering coherent the areas of overlap of the various behavioral disciplines. The analytical tools deployed in this task incorporate core principles from several behavioral disciplines. The proposed framework recognizes evolutionary theory, covering both genetic and cultural evolution, as the integrating principle of behavioral science. Moreover, if decision theory and game theory are broadened to encompass other-regarding preferences, they become capable of modeling all aspects of decision making, including those normally considered "psychological," "sociological" or "anthropological." The mind as a decision-making organ then becomes the organizing principle of psychology.

1 Introduction

The behavioral sciences include economics, biology, anthropology, sociology, psychology, and political science, as well as their subdisciplines, including neuroscience, archaeology and paleontology, and to a lesser extent, such related disciplines as history, legal studies, and philosophy.[1] These disciplines have many distinct concerns, but each includes a model of individual human behavior. These models are not only different, which is to be expected given their distinct explanatory goals, but *incompatible*. Nor can this incompatibility be accounted for by the type of causality involved (e.g., "ultimate" as opposed to "proximate" explanations). This situation is well known, but does not appear discomforting to behavioral scientists, as there has been virtually no effort to repair this condition.[2] In their current state,

---

[1] Biology straddles the natural and behavioral sciences. We include biological models of animal (including human) behavior, as well as the physiological bases of behavior, in the behavioral sciences.

however, according the behavioral sciences the status of true sciences is less than credible.

One of the great triumphs of Twentieth century science was the seamless integration of physics, chemistry, and astronomy, on the basis of a common model of fundamental particles and the structure of space-time. Of course, gravity and the other fundamental forces, which operate on extremely different energy scales, have yet to be reconciled, and physicists are often criticized for their seemingly endless generation of speculative models that might accomplish this reconciliation. But, a similar dissatisfaction with analytical incongruence on the part of their practitioners would serve the behavioral sciences well. This paper argues that we now have the analytical and empirical bases to construct the framework for an integrated behavioral science.

The behavioral sciences all include models of individual human behavior. These models should be compatible. Indeed, there should be a common underlying model, enriched in different ways to meet the particular needs of each discipline. We cannot easily attain this goal at present, however, as the various behavioral disciplines currently have *incompatible* models. Yet, recent theoretical and empirical developments have created the conditions for rendering coherent the areas of overlap of the various behavioral disciplines. The analytical tools deployed in this task incorporate core principles from several behavioral disciplines.[3]

The standard justification for the fragmentation of the behavioral disciplines is that each has a model of human behavior well suited to its particular object of study. While this is true, where these objects of study *overlap*, their models must be compatible. In particular, psychology, economics, anthropology, biology, and sociology should have concordant explanations of law-abiding behavior, charitable giving, political corruption, and voting behavior, and other complex behaviors that do not fit nicely within disciplinary boundaries. They do not.

This paper sketches a framework for the unification of the behavioral sciences. Two major conceptual categories, evolution and game theory, cover *ultimate* and *proximate* causality. Under each category are conceptual subcategories that relate to overlapping interests of two or more behavioral disciplines. I will argue the following points:

1. Evolutionary perspective: Evolutionary biology underlies all behavioral disciplines because *Homo sapiens* is an evolved species whose characteristics are the

---

[2] The last serious attempt at developing an analytical framework for the unification of the behavioral sciences was Parsons and Shils (1951). A more recent call for unity is Wilson (1998), which does not supply the unifying principles.

[3] A core contribution of political science, the concept of power, is absent from economic theory, yet interacts strongly with basic economic principles (Bowles and Gintis 2000). Lack of space prevents me from expanding on this important theme.

product of its particular evolutionary history.

1a. Gene-culture coevolution: The centrality of culture and complex social organization to the evolutionary success of *Homo sapiens* implies that fitness in humans will depend on the structure of cultural life.[4] Because culture is influenced by human genetic propensities, it follows that human cognitive, affective, and moral capacities are the product of a unique dynamic known as *gene-culture coevolution*, in which genes adapt to a fitness landscape of which cultural forms are a critical element, and the resulting genetic changes lay the basis for further cultural evolution. This coevolutionary process has endowed us with preferences that go beyond the self-regarding concerns emphasized in traditional economic and biological theories, and embrace such other-regarding values as a taste for cooperation, fairness, and retribution, the capacity to empathize, and the ability to value such constitutive behaviors as honesty, hard work, toleration of diversity, and loyalty to one's reference group.[5]

1b. Imitation and conformist transmission: Cultural transmission generally takes the form of *conformism*; that is, individuals accept the dominant cultural forms, ostensibly because it is fitness-enhancing to do so (Bandura 1977, Boyd and Richerson 1985, Conlisk 1988, Krueger and Funder 2004). Although adopting the beliefs, techniques, and cultural practices of successful individuals is a major mechanism of cultural transmission, there is constant cultural mutation, and individuals may adopt new cultural forms when those forms appear to better serve their interests (Gintis 1972, 2003a; Henrich 2001). One might expect that the analytical apparatus for understanding cultural transmission, including the evolution, diffusion, and extinction of cultural forms, might come from sociology or anthropology, the disciplines that focus on cultural life; but such is not the case. Both fields treat culture in a static manner that belies its dynamic and evolutionary character. By recognizing the common nature of genes and culture as forms of information that are transmitted intergenerationally, biology offers an analytical basis for understanding cultural transmission.

1c. Internalization of norms: In sharp contrast to other species, humans have preferences that are *socially programmable* in the sense that the individual's goals, and not merely the methods for their satisfaction, are acquired through a social learning process. Culture therefore takes the form not only of new techniques for controlling nature, but also of *norms and values* that are incorporated into individual preference functions through the sociological mechanism known as *socialization* and the psychological mechanism known as the *internalization of norms*. Surprisingly, the internalization of norms, which is perhaps the most singularly characteristic feature of the human mind, and central to understanding cooperation and conflict in human society, is ignored or misrepresented in the other behavioral disciplines, anthropology and social psychology aside.

2. Game theory: The analysis of living systems includes one concept that does not occur in the non-living world, and is not analytically represented in the natural sciences. This is the notion of a *strategic interaction*, in which the behavior of individuals is derived by assuming that each is choosing a *fitness-relevant response* to the actions of other individuals. The study of systems in which individuals choose fitness-relevant responses and in which such responses evolve dynamically, is called *evolutionary game theory*. Game theory provides a transdisciplinary conceptual basis for analyzing choice in the presence of strategic interaction. However, the classical game theoretic assumption that individuals are self-regarding must be abandoned except in specific situations (e.g. anonymous market interactions), and many characteristics that classical game theorists have considered logical implications from the principles of rational behavior, including the use of backward induction, are in fact not implied by rationality. Reliance on classical game theory has led economists and psychologists to mischaracterize many common human behaviors as irrational. Evolutionary game theory, whose equilibrium concept is that of a stable stationary point of a dynamical system, must therefore replace classical game theory, which erroneously favors subgame perfection and sequentiality as equilibrium concepts.

2a. The brain as a decision making organ: In any organism with a central nervous system, the brain evolved because centralized information processing enabled enhanced decision making capacity, the fitness benefits thereof more than offsetting its metabolic and other costs. Therefore, decision making must be the central organizing principle of psychology. This is not to say that learning (the focus of behavioral psychology) and information processing (the focus of cognitive psychology) are not of supreme importance, but rather that principles of learning and information processing only make sense in the context of the decision making role of the brain.[6]

2b. The rational actor model: General evolutionary principles suggest that individual decision making can be modeled as optimizing a preference function subject to informational and material constraints. Natural selection ensures that the content of preferences will reflect biological fitness, at least in the environments in which preferences evolved. The principle of expected utility extends this optimization to stochastic outcomes. The resulting model is called the *rational actor model* in economics, but I will generally refer to this as the *beliefs, preferences, and constraints* (BPC) model to avoid the often misleading connotations attached to the term "rational."[7]

Economics, biology and political science integrate game theory into the core of their models of human behavior. By contrast, game theory widely evokes emotions from laughter to hostility in the other behavioral disciplines. Certainly, if one rejects the BPC model (as these other disciplines characteristically do), game theory makes no sense whatever. The standard critiques of game theory in these other disciplines are indeed generally based on the sorts of arguments on which the critique of the BPC model are based, to which we turn in section 9.

In addition to these conceptual tools, the behavioral sciences of course share common access to the natural sciences, statistical and mathematical techniques, computer modeling, and a common scientific method.

The above principles are certainly not exhaustive; the list is quite spare, and will doubtless be expanded in the future. Note that I am not asserting that the above principles are the most important in each behavioral discipline. Rather, I am saying that they contribute to constructing a bridge across disciplines—a common model of human behavior from which each discipline can branch off.

Accepting the above framework may entail substantive reworking of basic theory in a particular discipline, but I expect that much research will be relatively unaffected by this reworking. For example, a psychologist working on visual processing, or an economist working on futures markets, or an anthropologist tracking food-sharing practices across social groups, or a sociologist gauging the effect of dual parenting on children's educational attainment, might gain little from knowing that a unified model underlay all the behavioral disciplines. But, I suggest that in such critical areas as the relationship between corruption and economic growth, community organization and substance abuse, taxation and public support for the welfare state, and the dynamics of criminality, researchers in one discipline are likely to benefit greatly from interacting with sister disciplines in developing valid and useful models.

In what follows, I will expand on each of the above concepts, after which I will address common objections to the beliefs, preferences, and constraints (BPC) model and game theory.

2 Evolutionary perspective

A *replicator* is a physical system capable of making copies of itself. Chemical crystals, such as salt, have this property of replication, 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 (Schrödinger called life an "aperiodic crystal" in 1944, before the structure of DNA was discovered), 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 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 *reduces* 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.

3 The Brain as a Decision Making Organ

The fitness of an organism depends on how effectively it make choices in an uncertain and varying environment. Effective choice must be a function of the organism's state of knowledge, which consists of the information supplied by the sensory organs that monitor the organism's internal states and its external environment. In relatively simple organisms, the choice environment is primitive and distributed in a decentralized manner over sensory inputs. But, in three separate groups of animals, the craniates (vertebrates and related creatures), arthropods (including insects, spiders, and crustaceans) and cephalopods (squid, octopuses, and other mollusks) a central nervous system with a brain (a centrally located decision making and control apparatus) evolved. The phylogenetic tree of vertebrates exhibits increasing complexity through time, and increasing metabolic and morphological costs of maintaining brain activity. *The brain evolved because more complex brains, despite their costs, enhanced the fitness of their bearers*. Brains therefore are ineluctably structured to make on balance fitness-enhancing decisions in the face of the various constellations of sensory inputs their bearers commonly experience.

The human brain shares most of its functions with that of other vertebrate species, including the coordination of movement, maintenance of homeostatic bodily functions, memory, attention, processing of sensory inputs, and elementary learning mechanisms. The distinguishing characteristic of the human brain, however, lies in its extraordinary power as a *decision making* mechanism.

Surprisingly, this basic insight is missing from psychology, which focuses on the processes that render decision-making possible (attention, logical inference, emotion vs. reason, categorization, relevance) but virtually ignores, and seriously misrepresents, decision-making itself. Psychology has two main branches: cognitive and behavioral. The former defines the brain as an "information-processing organ," and generally argues that humans are relatively poor, irrational, and inconsistent decision makers. The latter is preoccupied with learning mechanisms that humans share with virtually all metazoans (stimulus response, the law of effect, operant conditioning, and the like). For example, a widely used text of graduate-level readings in cognitive psychology, (Sternberg and Wagner 1999) devotes the 9th of 11 chapters to "Reasoning, Judgment, and Decision Making." It offers two papers, the first of which shows that human subjects generally fail simple logical inference tasks, and the second shows that human subjects are irrationally swayed by the way a problem is verbally "framed" by the experimenter. A leading undergraduate cognitive psychology text (Goldstein 2005) placed "Reasoning and Decision Making" the *last* of 12 chapters. It includes one paragraph describing the rational actor model, followed by many pages purporting to explain why the model is wrong. Behavioral psychology generally avoids positing internal states, of which beliefs and preferences, and even some constraints (e.g. such character virtues as keeping promises), are examples. When the rational actor model is mentioned with regard to human behavior, it is summarily rejected (Herrnstein, Laibson and Rachlin 1997). Not surprisingly, in a leading behavioral psychology text (Mazur 2002), choice is covered in the *last* of 14 chapters, and is limited to a review of the literature on choice between concurrent reinforcement schedules and the capacity to defer gratification.

Summing up a quarter century of psychological research in 1995, Paul Slovic asserted, accurately I believe, that "it is now generally recognized among psychologists that utility maximization provides only limited insight into the processes by which decisions are made." (Slovic 1995):365 "People are not logical," psychologists are fond of saying, "they are *psycho*logical." Of course, in this paper I argue precisely the opposite position: people are generally rational, though subject to performance errors.

Psychology could be the centerpiece of the human behavioral sciences by providing a general model of decision making for the other behavioral disciplines to use and elaborate for their various purposes. The field fails to hold this position because its core theories do not take the fitness-enhancing character of the human brain, its capacity to make effective decisions in complex environments, as central.[8]

4 The foundations of the BPC model

For every constellation of sensory inputs, each decision taken by an organism generates a probability distribution over fitness outcomes, the expected value of which

---

[8] The fact that psychology does not integrate the behavioral sciences is quite compatible, of course, with the fact that what psychologists do is of great scientific value.

is the *fitness* associated with that decision. Because fitness is a scalar variable (basically the expected number of offspring to reach reproductive maturity), for each constellation of sensory inputs, each possible action the organism might take has a specific fitness value; organisms whose decision mechanisms are optimized for this environment will choose the available action that maximizes this fitness value.[9] It follows that, given the state of its sensory inputs, if an orgasm 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. This is called *choice consistency*.

The so-called *rational actor model* was developed in the twentieth century by John von Neumann, Leonard Savage and many others. The model appears *prima facie* to apply only when individuals can determine all the logical and mathematical implications of the knowledge they possess. However, the model in fact depends only on choice consistency and the assumption that individuals can trade off among outcomes in the sense that for any finite set of outcomes $A_1, \ldots, A_n$, if A_1 is the least preferred and A_n the most preferred outcome, then for any A_i, $1 \leq i \leq n$ there is a probability p_i, $0 \leq p_i \leq 1$ such that the individual is indifferent between A_i and a lottery that pays A_1 with probability p_i and pays A_n with probability $1 - p_i$ (Kreps 1990). A lottery is a probability distribution over a finite set of monetary outcomes. Clearly, these assumptions are often extremely plausible. When applicable, the rational actor model's choice consistency assumption strongly enhances explanatory power, even in areas that have traditionally abjured the model (Coleman 1990, Kollock 1997, Hechter and Kanazawa 1997).

In short, when preferences are consistent, they can be represented by a numerical function, which we call the objective function, that individuals maximize subject to their beliefs (including Bayesian probabilities) and the constraints they face.

Four *caveats* are in order. First, this analysis does not suggest that people consciously maximize anything. Second, the model does *not* assume that individual choices, even if they are self-referring (e.g., personal consumption) are always welfare-enhancing. Third, preferences must be stable across time to be theoretically useful, but preferences are ineluctably a function of such parameters as hunger, fear, and recent social experience, and beliefs can change dramatically in response to immediate sensory experience. Finally, the BPC model does not presume that beliefs are correct or that they are updated correctly in the face of new evidence, although Bayesian assumptions concerning updating can be made part of preference consistency in elegant and compelling ways (Jaynes 2003).

---

[9] 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 (non-additive genetic) fitness has yet to be formally demonstrated, but the informal arguments are no less strong.

The rational actor model is the cornerstone of contemporary economic theory, and in the past few decades has become equally important in the biological modeling of animal behavior (Real 1991, Alcock 1993, Real and Caraco 1986). Economic and biological theory therefore have a natural affinity. The choice consistency on which the rational actor model of economic theory depends is rendered plausible by biological evolutionary theory, and the optimization techniques pioneered by economic theorists are routinely applied and extended by biologists in modeling the behavior of organisms.

For similar reasons, in a stochastic environment, natural selection will enhance the capacity of the brain to make choices that maximize expected fitness, and hence that satisfy the expected utility principle. To see this, suppose an organism must choose from action set X, where each $x \in X$ determines a lottery that pays i offspring with probability $p_i(x)$, for $i = 0, 1, \ldots, n$. Then the expected number of offspring from this lottery is

$$\psi(x) = \sum_{j=1}^{n} j p_j(x).$$

Let L be a lottery on X that delivers $x_i \in X$ with probability q_i for $i = 1, \ldots, k$. The probability of j offspring given L is then

$$\sum_{i=1}^{k} q_i p_j(x_i)$$

so the expected number of offspring given L is

$$\begin{aligned}\sum_{j=1}^{n} j \sum_{i=1}^{k} q_i p_j(x_i) &= \sum_{i=1}^{k} q_i \sum_{i=1}^{k} j p_j(x_i) \\ &= \sum_{i=1}^{k} q_i \psi(x_i),\end{aligned}$$

which is the expected value theorem with utility function $\psi(\cdot)$. See also Cooper (1987).

There are few reported failures of the expected utility theorem in non-humans, and there are some compelling examples of its satisfaction (Real and Caraco 1986). The difference between humans and other animals is that the latter are tested in *real life*, or in elaborate simulations of real life, whereas humans are tested in the laboratory under conditions differing radically from real life. Although it is important to know how humans choose in such situations (see section 9.7), there is certainly no guarantee they will make the same choices in the real-life situation that they make in the situation analytically generated to represent it. For example, a heuristic that says “adopt choice behavior that appears to have benefitted others” may lead to expected fitness or utility maximization even when individuals are error-prone when evaluating stochastic alternatives in the laboratory.

In addition to the explanatory success of theories based on the rational actor model, supporting evidence from contemporary neuroscience suggests that expected utility maximization is not simply an “as if” story. In fact, the brain’s neural circuitry makes choices by internally representing the payoffs of various alternatives as neural firing rates, choosing a maximal such rate (Glimcher 2003, Dorris and Bayer 2005). Neuroscientists increasingly find that an aggregate decision making process in the brain synthesizes all available information into a single, unitary value (Parker and Newsome 1998, Schall and Thompson 1999, Glimcher 2003). Indeed, when animals are tested in a repeated trial setting with variable reward, dopamine neurons appear to encode the difference between the reward that an animal expected to receive and the reward that an animal actually received on a particular trial (Schultz, Dayan and Montague 1997, Sutton and Barto 2000), an evaluation mechanism that enhances the environmental sensitivity of the animal’s decision making system. This error-prediction mechanism has the drawback of seeking only local optima (Sugrue, Corrado and Newsome 2005). Montague and Berns (2002) address this problem, showing that the obitofrontal cortex and striatum contain mechanisms for more global predictions that include risk assessment and discounting of future rewards. Their data suggest a decision making model that is analogous to the famous Black-Scholes options pricing equation (Black and Scholes 1973).

Although the neuroscientific evidence supports the BPC model, it does not support the traditional economic model of Homo economicus. For instance, recent evidence supplies a neurological basis for hyperbolic discounting, and hence undermines the traditional belief in time consistent preferences. For instance, McClure, Laibson, Loewenstein and Cohen (2004) showed that two separate systems are involved in long- vs. short-term decisions. The lateral prefrontal cortex and posterior parietal cortex are engaged in all intertemporal choices, while the paralimbic cortex and related parts of the limbic system kick in only when immediate rewards are available. Indeed, the relative engagement of the two systems is directly associated with the subject’s relative favoring of long- over short-term reward.

The BPC model is the most powerful analytical tool of the behavioral sciences. For most of its existence this model has been justified in terms of “revealed preferences,” rather than by the identification of neural processes that generate constrained optimal outcomes. The neuroscience evidence suggests a firmer foundation for the rational actor model.

5 Gene-Culture Coevolution

The genome encodes information that is used both to construct a new organism, to instruct the new organism how to transform sensory inputs into decision outputs (i.e., to endow the new organism with a specific preference structure), and to transmit this coded information virtually intact to the new organism. Because learning about one's environment may be costly and is error-prone, efficient information transmission will ensure that the genome encode all aspects of the organism's environment that are constant, or that change only very slowly through time and space. By contrast, environmental conditions that vary across generations and/or in the course of the organism's life history can be dealt with by providing the organism with the capacity to *learn*, and hence phenotypically adapt to specific environmental conditions.

There is an intermediate case that is not efficiently handled by either genetic encoding or 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 information by means other than the germ line concerning the current state of the environment. 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 lesser extent, in primates and other animals (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 influencing lower status), as in Henrich and Gil-White (2001), popularity-related as in Newman, Barabasi and Watts (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).[10] 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 (Lumsden and Wilson 1981, Cavalli-Sforza and Feldman 1982, Boyd

---

[10] For a more extensive analysis of the parallels between cultural and genetic evolution, see Mesoudi, Whiten and Laland (2006). I have borrowed heavily from that paper in this section.

and Richerson 1985).

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, Boyd and Richerson 1985). Moreover, there are strong interactions between genetic and epigenetic elements in human evolution, ranging from basic physiology (e.g., the transformation of the organs of speech with the evolution of language) to sophisticated social emotions, including empathy, shame, guilt, and guilt, and revenge-seeking (Zajonc 1980, 1984).

As a result of their common informational and evolutionary character, genetic and cultural modeling are strongly parallel (Mesoudi et al. 2006). Like biological transmission, cultural transmission occurs from parents to offspring, and like cultural transmission, which occurs horizontally between unrelated individuals, biological transmission in microbes and many plant species regularly transfers genes across lineage boundaries (Jablonka and Lamb 1995, Rivera and Lake 2004, Abbott, James, Milne and Gillies 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, archeologists who study cultural evolution have a *modus operandi* similar to that of paleobiologists who study genetic evolution (Mesoudi et al. 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 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, distinct, 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 over the past 35 years, have some of the fluidity of cultural units, whereas 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 undermine the standard view of the insular gene producing a single protein, and support the notion of genes having variable boundaries and 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, bee hives, 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*, is 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*— that is, a genetically induced environmental regularity becomes the basis for genetic selection, and genetic mutations that give rise to mutant niches survive if they are fitness enhancing for their constructors. The dynamical modeling of the reciprocal action of genes and culture is known as *gene-culture coevolution* (Lumsden and Wilson 1981, Durham 1991, Feldman and Zhivotovsky 1992, Bowles and Gintis 2005).

An excellent example of gene-environment coevolution is the honeybee, in which the origin of its eusociality doubtless lies in the high degree of relatedness fostered by haplodiploidy, but persists in modern species even though relatedness in the hive is generally quite low, as a result of multiple queen matings, multiple queens, queen deaths, and the like (Gadagkar 1991, Seeley 1997). The social structure of the hive is transmitted epigenetically across generations, and the honeybee 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). Although 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 both the neural plasticity of and the genetic basis for moral behavior. Brain regions involved in moral judgments and behavior include the prefrontal cortex, the orbitofrontal cortex, and the superior temporal sulcus (Moll, Zahn, di Oliveira-Souza, Krueger and Grafman 2005). These brain structures are present in all primates, but are 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 sociopathy (Miller, Darby, Benson, Cummings and Miller 1997) and the absence of embarrassment, pride and regret (Beer, Heerey, Keltner, Skabini and Knight 2003, Camille 2004).

6 The concept of culture across disciplines

Because of the centrality of culture to the behavioral sciences, it is worth noting the divergent use of the concept in distinct disciplines, and the sense in which it is used here.

Anthropology, the discipline that is most sensitive to the vast array of cultural groupings in human societies, treats culture as an expressive totality defining the life space of individuals, including symbols, language, beliefs, rituals, and values.

By contrast, in biology culture is generally treated as *information*, in the form of instrumental techniques and practices, such as those used in producing of necessities, fabricating tools, waging war, defending territory, maintaining health, and rearing children. We may include in this category "conventions" (e.g., standard greetings, forms of dress, rules governing the division of labor, the regulation of marriage, and rituals) that differ across groups and serve to coordinate group behavior, facilitate communication and maintain shared understandings. Similarly, we may include *transcendental beliefs* (e.g., that sickness is caused by angering the gods, that good deeds are rewarded in the afterlife) as a form of information. A transcendental belief is the assertion of a state of affairs that has a truth value, but one that believers either cannot or choose not to test personally (Atran 2004). Cultural transmission in humans, in this view, is therefore a process of information transmission, rendered possible by our uniquely prodigious cognitive capacities (Tomasello, Carpenter, Call, Behne and Moll 2005).

The predisposition of a new member to accept the dominant cultural forms of a group is called *conformist transmission* (Boyd and Richerson 1985). Conformist transmission may be fitness enhancing because, if an individual must determine the most effective of several alternative techniques or practices, and if experimentation is costly, it may be payoff-maximizing to copy others rather than incur the costs of experimenting (Boyd and Richerson 1985, Conlisk 1988). Conformist transmission extends to the transmission of transcendental beliefs as well. Such beliefs affirm techniques where the cost of experimentation is extremely high or infinite, and the cost of making errors is high as well. This is, in effect, Blaise Pascal's argument for the belief in God. This view of religion is supported by Boyer (2001), who models transcendental beliefs as cognitive beliefs that coexist and interact with our other more mundane beliefs. In this view, one conforms to transcendental beliefs because their truth value has been ascertained by others (relatives, ancestors, prophets), and are deemed to be as worthy of affirmation as the everyday techniques and practices, such as norms of personal hygiene, that one accepts on faith, without personal verification.

Sociology and anthropology recognize the importance of conformist transmission, but the notion is virtually absent from economic theory. For example, in economic theory consumers maximize utility and firms maximize profits by considering only market prices and their own preference and production functions. In fact, in the face of incomplete information and the high cost of information-gathering, both consumers and firms in the first instance may simply imitate what appear to be the successful practices of others, adjust their behavior incrementally in the face of varying market conditions, and sporadically inspect alternative strategies in limited areas (Gintis 2004).

Possibly part of the reason the BPC model is so widely rejected in some disciplines is the belief that optimization is analytically incompatible with reliance on imitation and hence with conformist transmission. In fact, the economists' distaste for optimization *via* imitation is not complete (Conlisk 1988, Bikhchandani, Hirshleifer and Welsh 1992), and it is simply a doctrinal prejudice. Recognizing that imitation is an aspect of optimization has the added attractiveness of allowing us to model cultural change in a dynamic manner: as new cultural forms displace older forms when they appear to advance the goals of their bearers (Henrich 1997, Henrich and Boyd 1998, Henrich 2001, Gintis 2003a).

7 Programmable preferences and the sociology of choice

Sociology, in contrast to biology, treats culture primarily as a set of *moral values* (e.g., norms of fairness, reciprocity, justice) that are held in common by members of the community (or a stratum within the community) and are transmitted from generation to generation by the process of *socialization*. According to Durkheim (1951), the organization of society involves assigning individuals to specific *roles*, each with its own set of socially sanctioned values. A key tenet of socialization theory is that a society's values are passed from generation to generation through the *internalization of norms* (Durkheim 1951, Benedict 1934, Mead 1963, Parsons 1967, Grusec and Kuczynski 1997, Nisbett and Cohen 1996, Rozin, Lowery, Imada and Haidt 1999), which is 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 contrast with anthropology and biology could hardly be more complete. Unlike anthropology, which celebrates the irreducible heterogeneity of cultures, sociology sees cultures as sharing much in common throughout the world (Brown 1991). In virtually every society, says sociology, youth are pressed to internalize the value of being trustworthy, loyal, helpful, friendly, courteous, kind, obedient, cheerful, thrifty, brave, clean, and reverent (famously captured by the Boy Scouts of America). In biology, values are collapsed into techniques and the machinery of internalization is unrepresented.

Internalized norms are followed not because of their epistemic truth value, but because of their moral value. In the language of the BPC model, internalized norms are accepted not as instruments towards achieving other ends, but rather as *arguments in the preference function that the individual maximizes*, or are *self-imposed constraints*. For example, individuals who have internalized the value of "speaking truthfully" will constrain themselves to do so even in some cases where the net payoff to speaking truthfully would otherwise be negative. Internalized norms are therefore *constitutive* in the sense that an individual strives to live up to them*for their own sake*. Fairness, honesty, trustworthiness, and loyalty are ends, not means, and 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.

The human responsiveness to socialization pressures represents the most powerful form of epigenetic transmission found in nature. In effect, *human preferences are programmable*, in the same sense that a digital computer can be programmed to perform a wide variety of tasks. This epigenetic flexibility, which is an emergent property of the complex human brain, in considerable part accounts for the stunning success of the species *Homo sapiens*. When people internalize a norm, the frequency of its occurrence in the population will be higher than if people follow the norm only instrumentally—i.e., only when they perceive it to be in their material self-interest to do so. The increased incidence of altruistic prosocial behaviors permits humans to cooperate effectively in groups (Gintis, Bowles, Boyd and Fehr 2005).

Given the abiding disarray in the behavioral sciences, it should not be surprising to find that socialization has no conceptual standing outside of sociology, anthropology, and social psychology, and that most behavioral scientists subsume it under the general category of "information transmission," which would make sense only if moral values expressed matters of fact, which they do not. Moreover, the socialization concept is incompatible with the assumption in economic theory that preferences are mostly, if not exclusively, self-regarding, given that social values commonly involve caring about fairness and the well-being of others. Sociology, in turn, systematically ignores the limits to socialization (Tooby and Cosmides 1992, Pinker 2002) and supplies no theory of the emergence and abandonment of particular values, both of which in fact depend in part on the contribution of the values to fitness and well-being, as economic and biological theory would suggest (Gintis 2003a,b). Moreover, there are often swift society-wide value changes that cannot be accounted for by socialization theory (Wrong 1961, Gintis 1975). When properly qualified, however, and appropriately related to the general theory of cultural evolution and strategic learning, socialization theory is considerably strengthened.

8 Game theory: the universal lexicon of life

In the BPC model, choices give rise to probability distributions over outcomes, the expected values of which are the payoffs to the choice from which they arose. Game theory extends this analysis to cases where there are multiple decision makers. In the language of game theory, *players* (or *agents*) are endowed with a set of *strategies*, they have certain *information* concerning the rules of the game, the nature of the other players and their available strategies. Finally, for each combination of strategy choices by the players, the game specifies a distribution of *individual payoffs* to the players. Game theory predicts the behavior of the players by assuming each maximizes its preference function subject to its information, beliefs, and constraints (Kreps 1990).

Game theory is a logical extension of evolutionary theory. To see this, suppose there is only one replicator, deriving its nutrients and energy from non-living sources (the sun, the Earth's core, amino acids produced by electrical discharge, and the like). The replicator population will then grow at a geometric rate, until it presses on its environmental inputs. At that point, mutants that exploit the environment more efficiently will out-compete their less efficient conspecifics, and with input scarcity, mutants will emerge that "steal" from conspecifics that have amassed valuable resources. With the rapid growth of such mutant predators, their prey will mutate, thereby devising means of avoiding predation, and the predators will counter with their own novel predatory capacities. In this manner, strategic interaction is born from elemental evolutionary forces. It is only a conceptually short step from this point to cooperation and competition among cells in a multi-cellular 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 (Maynard Smith and Szathmary 1997).

Historically, game theory emerged not from biological considerations, but rather from the strategic concerns of combatants 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 individuals possessed of formidable reasoning and information processing capacity. Developments within game theory in recent years, however, render this caricature inaccurate.

First, game theory has become the basic framework for modeling animal behavior (Maynard Smith 1982, Alcock 1993, Krebs and Davies 1997), and as a result has shed its static and hyperrationalistic character, in the form of evolutionary game theory (Gintis 2000a). Evolutionary game theory does 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 evolutionary game-theoretic models (Young 1998, Gintis 2000a, Gigerenzer and Selten 2001). Therefore, we may model individuals as considering only a restricted subset of strategies (Winter 1971, Simon 1972), and as using rule-of-thumb heuristics rather than maximization techniques (Gigerenzer and Selten 2001). Game theory is therefore a generalized schema that permits the precise framing of meaningful empirical assertions, but imposes no particular structure on the predicted behavior.

Second, evolutionary game theory has become key to understanding the most fundamental principles of evolutionary biology. 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—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 genetic 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 have been generalized by Liberman (1988), Hammerstein and Selten (1994), Hammerstein (1996), Eshel, Feldman and Bergman (1998) and others.

Third, the most natural setting for biological and social 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 strategy is to be deployed in reaction 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, perhaps in part through the action of *mirror neurons*, which fire when either the individual performs a task or undergoes an experience, or when the individual observes another individual performing the same task or undergoing the same experience (Williams, Whiten, Suddendorf and Perrett 2001, Rizzolatti, Fadiga, Fogassi and Gallese 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 (Taylor and Jonker 1978).

Fourth, behavioral game theorists, who used game theory to collect experimental data concerning strategic interaction, now widely recognize that in many social interactions, individuals are not self-regarding. Rather, they often care about the payoffs to and intentions of other players, and will sacrifice to uphold personal standards of honesty and decency (Fehr and Gächter 2002, Wood 2003, Gintis et al. 2005, Gneezy 2005). Moreover, humans care about power, self-esteem, and behaving morally (Gintis 2003b, Bowles and Gintis 2005, Wood 2003). Because the rational actor model treats action as instrumental towards achieving rewards, it is often inferred that action itself cannot have reward value. This is an unwarranted inference. For example, the rational actor model can be used to explain collective action (Olson 1965), because individuals may place positive value on the process of acquisition (e.g., "fighting for one's rights"), and they can value punishing those who refuse to join in the collective action (Moore, Jr. 1978, Wood 2003). Indeed, contemporary experimental work indicates that one can apply standard choice theory, including deriving of demand curves, plotting concave indifference curves, and finding price elasticities, for such preferences as charitable giving and punitive retribution (Andreoni and Miller 2002).

As a result of its maturation over the past quarter century, game theory is well positioned to serve as a bridge across the behavioral sciences, providing both a lexicon for communicating across fields with distinct and incompatible conceptual systems, and a theoretical tool for formulating a model of human choice that can serve all the behavioral disciplines.

9 Some misconceptions concerning the BPC model and game theory

Many behavioral scientists reject the BPC model and game theory on the basis of one or more of the following arguments. In each case, I shall indicate why the objection is not compelling.

9.1 Individuals are only boundedly rational

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 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 the Second Law of Thermodynamics, 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 example, 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. As we have seen, as long as individuals have consistent preferences, they can be modeled as maximizing an objective function. Of course, if there is a precise objective (e.g., solve the problem with an exogenously given degree of accuracy), then the information contained in knowledge of preference consistency may be ignored. But, once the degree of accuracy is treated as endogenous, multiple objectives compete (e.g., cost and accuracy), and the BPC model cannot be ignored. 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.

9.2 Decision makers are not consistent

It is widely argued that in many situations of extreme importance choice consistency fails, so preferences are not maximized. These cases include time inconsistency, in which individuals have very high short-term discount rates and much lower long-term discount rates (Herrnstein 1961, Ainslie 1975, Laibson 1997). As a result, people lack the will-power to sacrifice present pleasures for future well-being. This leads to such well-known behavioral problems as unsafe sex, crime, substance abuse, procrastination, under-saving, and obesity. It is thus held that these phenomena of great public policy importance are irrational and cannot be treated with the BPC model.

When the choice space for time preference consists of pairs of the form (*re-*

ward,delay until reward materializes), then preferences are indeed time inconsistent. The long-term discount rate can be estimated empirically at about 3% per year (Huang and Litzenberger 1988, Rogers 1994), but short-term discount rates are often an order of magnitude or more greater than this (Laibson 1997). Animal studies find rates are several orders of magnitude higher (Stephens, McLinn and Stevens 2002). Consonant with these findings, sociological theory stresses that *impulse control*—learning to favor long-term over short-term gains—is a major component in the socialization of youth (Mischel 1974, Power and Chapieski 1986, Grusec and Kuczynski 1997).

However, suppose we expand the choice space to consist of triples of the form (*reward,current time,time when reward accrues*), for example, so that $(\pi_1, t_1, s_1) > (\pi_2, t_2, s_2)$ means that at the individual prefers to be at time t_1 facing a reward π_1 delivered at time s_1 to being at time t_2 facing a reward π_2 delivered at time s_2. Then the observed behavior of individuals with discount rates that decline with the delay become choice consistent, and there are two simple models that are roughly consistent with the available evidence (and differ only marginally from each other): hyperbolic and quasi-hyperbolic discounting (Fishburn and Rubinstein 1982, Ainslie and Haslam 1992, Ahlbrecht and Weber 1995, Laibson 1997). The resulting BPC models allow for sophisticated and compelling economic analyses of policy alternatives (Laibson, Choi and Madrian 2004).

Other observed instances of *prima facie* choice inconsistency can be handled in a similar fashion. For example, in experimental settings, individuals exhibit *status quo* bias, loss aversion, and regret—all of which imply inconsistent choices (Kahneman and Tversky 1979, Sugden 1993). In each case, however, choices become consistent by a simple redefinition of the appropriate choice space. Kahneman and Tversky's "prospect theory," which models *status quo* bias and loss aversion, is precisely of this form. Gintis (2006) has shown that this phenomenon has an evolutionary basis in territoriality in animals and in pre-institutional property rights in humans.

There remains perhaps the most widely recognized example of inconsistency, that of 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 the lottery choice and toward the higher maximum amount of winnings in monetary valuation. If this were true for lotteries in general it might compromise the BPC model.[11] 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 example, in Grether and Plott (1979) the average difference between expected values of comparison pairs was 2.51% (calculated from their Table 2, p. 629). The corresponding figure for Tversky et al. (1990) was 13.01%. When the choices involve small amounts of money and are so close to equal expected value, it is not surprising that inappropriate cues are relied upon to determine choice. Moreover, Berg, Dickhaut and Rietz (2005) have shown that when analysis is limited to studies that have truth-revealing incentives, preference reversals are well described by a model of maximization with error.

Another source of inconsistency is that observed preferences may not lead to the well-being, or even the immediate pleasure, of the decision maker. For example, fatty foods and tobacco injure health yet are highly prized, addicts often say they get no pleasure from consuming their drug of choice, but are driven by an inner compulsion to consume, and individuals with obsessive-compulsive disorders repeatedly perform actions that they know are irrational and harmful. More generally, behaviors resulting from excessively high short-term discount rates, discussed above, are likely to lead to a divergence of choice and welfare.

However, the BPC model is based on the premise that choices are consistent, not that choices are highly correlated with welfare. Drug addiction, unsafe sex, unhealthy diet, and other individually welfare-reducing behaviors can be analyzed with the BPC model, although in such cases preferences and welfare may diverge. I have argued that we can expect the BPC to hold because, on an evolutionary time scale, brain characteristics will be selected according to their capacity to contribute to the fitness of their bearers. But, fitness cannot be equated with well-being in any creature. Humans, in particular, live in an environment so dramatically different from that in which our preferences evolved that it seems to be miraculous that we are as capable as we are of achieving high levels of individual well-being. For example, in virtually all known cases, fertility increases with per capital material wealth in a society up to a certain point, and then decreases. This is known as the *demographic transition*, and accounts for our capacity to take out increased technological power in the form of consumption and leisure rather than increased numbers of offspring (Borgerhoff Mulder 1998). No other known creature behaves in this fashion. Therefore, our preference predispositions have not "caught up" with

---

[11] I say "might" because in real life individuals generally do not choose among lotteries by observing or contemplating probabilities and their associated payoffs, but by imitating the behavior of others who appear to be successful in their daily pursuits. In frequently repeated lotteries, the Law of Large Numbers ensures that the higher expected value lottery will increase in popularity by imitation without any calculation by participants.

our current environment and, especially given the demographic transition and our excessive present-orientation, they may never catch up (Elster 1979, Akerlof 1991, O'Donoghue and Rabin 2001).

9.3 Addiction contradicts the BPC model

Substance abuse is of great contemporary social importance and appears most clearly to violate the notion of rational behavior. Substance abusers are often exhibited as prime examples of time inconsistency and the discrepancy between choice and well-being, but as discussed above, these characteristics do not invalidate the use of the BPC model. More telling, perhaps, is the fact that even draconian increases in the penalties for illicit substance use do not lead to the abandonment of illegal substances. In the United States, for example, the "war on drugs" has continued for several decades, despite dramatically increasing the prison population, it has not effectively curbed the illicit behavior. Because the hallmark of the rational actor model is that individuals trade off among desired goals, the lack of responsiveness of substance abuse to dramatically increased penalties has led many researchers to reject the BPC model out of hand.

The target of much of the criticism of the rational actor approach to substance abuse is the work of economist Gary Becker and his associates, in particular, the seminal paper Becker and Murphy (1988). Many aspects of the Becker-Murphy "rational addiction" model are accurate, however, and subsequent empirical research has strongly validated the notion that illicit drugs respond to market forces much as any marketed good or service. For example 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. The price elasticities for heroin and cocaine were about 1.70 and 0.96, respectively, which are quite high. Using these figures, the authors estimate that the 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 Koszegi (2001) explain this by presenting evidence that drug users exhibit the commitment and self-control problems that are typical of time-inconsistent individuals, for whom the possible future penalties have highly attenuated deterrent value in the present. Nevertheless, allowing for this attenuated value, sophisticated economic analysis, of the sort developed by Becker, Grossman and Murphy (1994) can be deployed for policy purposes. Moreover, this analytical and quantitative analysis harmonizes with the finding that, along with raising the price of cigarettes, the most effective way to reduce the incidence of smoking is to raise its immediate personal costs, such as being socially stigmatized, being banned from smoking in public buildings, and being considered impolite, given the well-known externalities associated with second-hand smoke (Brigden and De Beyer 2003).

9.4 Positing exotic tastes explains nothing

Broadening the rational actor model beyond its traditional form in neoclassical economics runs the risk of developing unverifiable and *post hoc* theories, as our ability to theorize outpaces our ability to test theories. Indeed, the folklore among economists dating back at least to Becker and Stigler (1977) is that "you can always explain any bizarre behavior by assuming sufficiently exotic preferences."

This critique was telling before researchers had the capability of actually measuring preferences and testing the cogency of models with nonstandard preferences (i.e., preferences over things other than marketable commodities, forms of labor, and leisure). However, behavioral game theory now provides the methodological instruments for devising experimental techniques that allow us to estimate preferences with some degree of accuracy, (Gintis 2000a, Camerer 2003). Moreover, we often find that the appropriate experimental design variations can generate novel data allowing us to distinguish among models that are equally powerful in explaining the existing data (Tversky and Kahneman 1981, Kiyonari, Tanida and Yamagishi 2000). Finally, because behavioral game-theoretic predictions can be systematically tested, the results can be replicated by different laboratories (Plott 1979, V. Smith 1982, Sally, 1995), and models with very few nonstandard preference parameters, examples of which are provided in Section 10 below, can be used to explain a variety of observed choice behavior,

9.5 Decisions are sensitive to framing bias

The BPC model assumes that individuals have stable preferences and beliefs that are functions of the individual's personality and current needs. Yet, in many cases laboratory experiments show that individuals can be induced to make choices over payoffs based on subtle or obvious cues that ostensibly do not affect the value of the payoffs to the decision maker. For example, if a subjects' partner in an experimental game is described as an "opponent," or the game itself is described as a "bargaining game," subjects may make very different choices than when the partner is described as a "teammate", or the game is described as a community participation game. Similarly, a subject in an experimental game may reject an offer if made by his bargaining partner, but accept the same offer if made by the random draw of a computer on behalf of the proposer (Blount 1995).

Sensitive to this critique, experimenters in the early years of behavioral game theory attempted to minimize the possibility of framing effects by rendering as abstract and unemotive as possible the language in which a decision problem or strategic interaction was described. It is now widely recognize that it is in fact impossible to avoid framing effects, because abstraction and lack of real-world reference are themselves a frame rather than an absence thereof. A more productive way to deal with framing is to make the frame a part of the specification of the experiment itself. Varying the frame systematically will uncover the effect of the frame on the choices of the subjects, and by inference, on their beliefs and preferences.

We do not have a complete understanding of framing, but we know enough to assert that its existence does not undermine the BPC model. If subjects care only about the "official" payoffs in a game, and if framing does not affect the beliefs of the subjects as to what other subjects will do, then framing could not affect behavior in the BPC framework. But, subjects generally do care about fairness, reciprocity, and justice as well as about the game's official payoffs, and when confronted with a novel social setting in the laboratory, subjects must first decide what moral values to apply to the situation by *mapping the game onto some sphere of everyday life* to which they are accustomed. The verbal and other cues provided by experimenters are the clues that subjects use to "locate" the interaction in their social space, so that moral principles can be properly applied to the novel situation. Moreover, framing instruments such as calling subjects "partners" rather than "opponents" in describing the game can increase cooperation because strong reciprocators (Gintis 2000b), who prefer to cooperate if others do the same, may increase their assessment of the probability that others will cooperate (see section 10), given the "partner" as opposed to the "opponent" cue. In sum, framing is in fact an ineluctable part of the BPC model, properly construed.

9.6 People are faulty logicians

The BPC model permits us to infer the beliefs and preferences of individuals from their choices under varying constraints. Such inferences are valid, however, only if individuals can intelligently vary their behavior in response to novel conditions. While it is common for behavioral scientists who reject the BPC model to explain an observed behavior as the result of an error or confusion on the part of the individual, the BPC model is less tolerant of such explanations if individuals are reasonably well-informed and the choice setting reasonably transparent and easily analyzable.

Evidence from experimental psychology over the past 40 years has led some psychologists to doubt the capacity of individuals to reason sufficiently accurately to warrant the BPC presumption of subject intelligence. For example, in one well-known experiment performed by Tversky and Kanheman (1983), a young woman Linda is described as politically active in college and highly intelligent, then the subject is asked which of the following two statements is more likely: "Linda is a bank teller" or "Linda is a bank teller and is active in the feminist movement." Many subjects rate the second statement more likely, despite the fact that elementary probability theory asserts that if p implies q, then p cannot be more likely than q. Because the second statement implies the first, it cannot be more likely than the first.

I personally know many people (though not scientists) who give this "incorrect" answer, and I never have observed these individuals making simple logical errors in daily life. Indeed, in the literature on the "Linda problem" several alternatives to faulty reasoning have been offered. One highly compelling alternative is based on the notion that in normal conversation, a listener assumes that any information provided by the speaker is relevant to the speaker's message (Grice 1975). Applied to this case, the norms of conversation lead the subject to believe that the experimenter wants Linda's politically active past to be taken adequately into account (Hilton 1995, Wetherick 1995). Moreover, the meaning of such terms as "more likely" or "higher probability" are vigorously disputed even in the theoretical literature, and hence are likely to have a different meaning for the average subject versus for the expert. For example, if I were given two piles of identity folders and ask to search through them to find the one belonging to Linda, and one of the piles was "all bank tellers" while the other was "all bank tellers who are active in the feminist movement," I would surely look through the second (doubtless much smaller) pile first, even though I am well aware that there is a "higher probability" that Linda's folder is in the former pile rather than the latter one.

More generally, subjects may appear irrational because basic terms have different meanings in propositional logic versus in everyday logical inference. For example, "if p then q" is true in formal logic except when p is true and q is false. In everyday usage "if p then q" may be interpreted as a material implication, in which there is something about p that causes q to be the case. In particular, in material logic "p implies q" means "p is true and this situation causes q to be true." Similarly, "if France is in Africa, then Paris is in Europe" is true in propositional logic, but false as a material implication. Part of the problem is also that individuals without extensive academic training simply lack the expertise to follow complex chains of logic, so psychology experiments often exhibit a high level of *performance error* (Cohen 1981; see section 11). For example, suppose Pat and Kim live in a certain town where all men have beards and all women wear dresses. Then the following can be shown to be true in propositional logic: "Either if Pat is a man then Kim wears a dress or if Kim is a woman, then Pat has a beard." It is quite hard to see why this is formally, true, and it is not true if the implications are material. Finally, the logical meaning of "if p then q" can be context dependent. For example, "if you eat dinner (p), you may go out to play (q)" formally means "you may go out to play (q) only if you eat dinner (p)."

We may apply this insight to an important strand of experimental psychology that purports to have shown that subjects systematically deviate from simple principles of logical reasoning. In a widely replicated study, Wason (1966) showed subjects cards each of which had a "1" or "2" on one side and "A" or "B" on the other, and stated the following rule: a card with a vowel on one side must have an odd number on the other. The experimenter then showed each subject four cards, one showing "1", one showing "2", one showing "A", and one showing "B", and asked the subject which cards must be turned over to check whether the rule was followed. Typically, only about 15% of college students point out the correct cards ("A" and "2"). Subsequent research showed that when the problem is posed in more concrete terms, such as "any person drinking beer must be more than 18," the correct response rate increases considerably (Cheng and Holyoak 1985, Cosmides 1989, Stanovich 1999, Shafir and LeBoeuf 2002). This accords with the observation that most individuals do not appear to have difficulty making and understanding logical arguments in everyday life.

9.7 People are poor statistical decision makers

Just as the rational actor model began to take hold in the mid-Twentieth century, vigorous empirical objections began to surface. The first was Allais (1953), who described cases in which subjects exhibited clear inconsistency in choosing among simple lotteries. It has been shown that Allais' examples can be explained by regret theory (Bell 1982, Loomes and Sugden 1982), which can be represented by consistent choices over pairs of lotteries (Sugden 1993).

Close behind Allais came the famous Ellsberg Paradox (Ellsberg 1961), which can be shown to violate the most basic axioms of choice under uncertainty. Consider two urns. Urn A has 51 red balls and 49 white balls. Urn B also has 100 red and white balls, but the fraction of red balls is unknown. Subjects are asked to choose in two situations. In each, the experimenter draws one ball from each urn but the two balls remain hidden from the subject's sight. In the first situation, the subject can choose the ball that was drawn from urn A or urn B, and if the ball is red, the subject wins $10. In the second situation, the subject again can choose the ball drawn from urn A or urn B, and if the ball is white, the subject wins $10. Many subjects choose the ball drawn from urn A in both situations. This obviously violates the expected utility principle, no matter what probability the subject places on the probability the ball from urn B is white.

It is easy to see why unsophisticated subjects make this error. Urn B seems to be *riskier* than urn A, because we know the probabilities in A but not in B. It takes a relatively sophisticated probabilistic argument—one that no human being ever made or could have made (to our knowledge) prior to the modern era—to see that in fact in this case uncertainty does not lead to increased risk. Indeed, most intelligent subjects who make the Ellsberg error will be convinced, when presented with the logical analysis, to modify their choices without modifying their preferences. In cases like this, we speak of performance error, whereas in cases such as the Allais Paradox, even the most highly sophisticated subject will need to change his choice unless convinced to change his preference ordering.

Numerous experiments document that many people have beliefs concerning probabilistic events that are without scientific foundation, and which will likely lead them to sustain losses if acted upon. For example, virtually every enthusiast believes that athletes in competitive sports run “hot and cold,” although this has never been substantiated empirically. In basketball, when a player has a “hot hand,” he is preferentially allowed to shoot again, and when he has a “cold hand,” he is often taken out of the game. I have yet to meet a basketball fan who does not believe in the phenomenon of the hot hand. Yet, Gilovich, Vallone and Tversky (1985) have shown on the basis of a statistical analysis using professional basketball data, that the hot hand does not exist.[12] This is but one instance of the general rule that our brains often lead us to perceive a pattern when faced with purely random data. In the same vein, I have talked to professional stock traders who believe, on the basis of direct observation of stock volatility, that stocks follow certain laws of inertia and elasticity that cannot be found through a statistical analysis of the data. Another example of this type is the “gambler’s fallacy,” which is that in a fair game, the appearance of one outcome several times in a row renders that outcome less likely in the next several plays of the game. Those who believe this cannot be dissuaded by scientific evidence. Many who believe in the “Law of Small Numbers,” which says that a small sample from a large population will have the same distribution of characteristics as the population (Tversky and Kanheman 1971), simply cannot be dissuaded either by logical reasoning or presentation of empirical evidence.

We are indebted to Daniel Kahneman, Amos Tversky, and their colleagues for a long series of brilliant papers, beginning in the early 1970’s, documenting the various errors intelligent subjects commit in dealing with probabilistic decision making. Subjects systematically underweight base rate information in favor of

---

[12] I once presented this evidence to graduating seniors in economics and psychology at Columbia University, towards the end of a course that developed and used quite sophisticated probabilistic modeling. Many indicated in their essays that they did not believe the data.

salient and personal examples, they reverse lottery choices when the same lottery is described by emphasizing probabilities rather than monetary payoffs, or when described in term of losses from a high baseline as opposed to gains from a low baseline, and they treat proactive decisions differently from passive decisions even when the outcomes are exactly the same, and when outcomes are described in terms of probabilities as opposed to frequencies (Kahneman, Slovic and Tversky 1982, Kahneman and Tversky 2000).

These findings are important for understanding human decision making and for formulating effective social policy mechanisms where complex statistical decisions must be made. However, these findings are not a threat to the BPC model (Gigerenzer and Selten 2001). They are simply performance errors in the form of incorrect beliefs as to how payoffs can be maximized.[13]

Statistical decision theory did not exist until recently. Before the contributions of Bernoulli, Savage, von Neumann and other experts, no creature on Earth knew how to value a lottery. It takes years of study to feel at home with the laws of probability. Moreover, it is costly, in terms of time and effort, to apply these laws even if we know them. Of course, if the stakes are high enough, it is worthwhile to go to the effort, or engage an expert who will do it for you. But generally, we apply a set of heuristics that more or less get the job done (Gigerenzer and Selten 2001). Among the most prominent heuristics is simply *imitation*: decide what class of phenomenon is involved, find out what people "normally do" in that situation, and do it. If there is some mechanism leading to the survival and growth of relatively successful behaviors and if the problem in question recurs with sufficient regularity, the choice-theoretic solution will describe the winner of a dynamic social process of trial, error, and replication through imitation.

9.8 Classical game theory misunderstands rationality

Game theory predicts that rational agents will play Nash equilibria. Because my proposed framework includes both game theory and rational agents, I must address that fact that in important cases, the game theoretic prediction is ostensibly falsified by the empirical evidence. The majority of examples of this kind arise from the

---

[13] In a careful review of the field, Shafir and LeBoeuf (2002) reject the performance error interpretation of these results, calling this a "trivialization" of the findings. They come to this conclusion by asserting that performance errors must be randomly distributed, whereas the errors found in the literature are systematic and reproducible. These authors, however, are mistaken in believing that performance errors must be random. Ignoring base rates in evaluating probabilities or finding risk in the Ellsberg two urn problems are surely performance errors, but the errors are quite systematic. Similarly, folk intuitions concerning probability theory lead to highly reproducible results, although incorrect.

assumption that individuals are self-regarding, which can be dropped without violating the principles of game theory. Game theory also offers solutions to problems of cooperation and coordination that are never found in real life, but in this case, the reason is that the game theorists assume perfect information, the absence of errors, the use of solution concepts that lack plausible dynamical stability properties, or other artifices without which the proposed solution would not work (Gintis 2005). However, in many cases, rational individuals simply do not play Nash equilibria at all under plausible conditions.

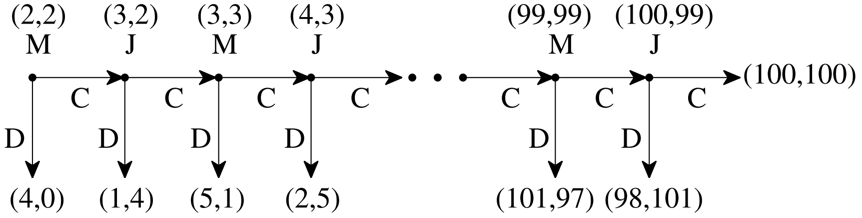

Figure 1: The Hundred Round Centipede Game illustrates the fallacy of holding that "rational" agents must use backward induction in their strategic interaction.

Consider, for example, the centipede game, depicted in Figure 1 (Rosenthal 1981, Binmore 1987). It is easy to show that this game has only one Nash payoff structure, in which player one defects on round one. However, when people actually play this game, they generally cooperate until the last few rounds (McKelvey and Palfrey 1992). Game theorists are quick to call such cooperation "irrational." For example, Reinhard Selten (himself a strong supporter of "bounded rationality") considers any move other that immediate defection a "failure to behave according to one's rational insights" (Selten 1993):133. This opinion is a result of the fact that this is the unique Nash equilibrium to the game, it does not involve the use of mixed strategies, and it can be derived from backward induction. However, as the professional literature makes abundantly clear, it is simply not true that rational agents must use backward induction. Rather, the most that rationality can ensure is *rationalizability* (Bernheim 1984, Pearce 1984), which in the case of the centipede game includes any pair of actions, except for cooperation on a player's final move. . Indeed, the epistemic conditions under which it is reasonable to assert that rational agents will play a Nash equilibrium are plausible in only the simplest cases (Aumann and Brandenburger 1995).

Another way to approach this issue is to begin by simply endowing each player with a BPC structure, and defining each player's "type" to be the round on which the player would first defect, assuming this round is reached. The belief system of each player is then a subjective probability distribution over the type of his opponent. It is clear that if players attempt to maximize their payoffs subject to this probability distribution, many different actions can result. Indeed, when people play this game, they generally cooperate at least until the final few rounds. This, moreover, is an eminently the correct solution to the problem, and much more lucrative that not the Nash equilibrium. Of course, one could argue that both players must have the same subjective probability distribution (this is called the *common priors* assumption), in which case (assuming common priors are common knowledge) there is only one equilibrium, the Nash equilibrium. But, it is hardly plausible to assume two players have the same subjective probability distribution over the types of their opponents without giving a mechanism that would produce this result.[14] In a famous paper Nobel prize winning economist John Harsanyi (1967) argued that common priors follow from the assumption that individuals are rational, but this argument depends on a notion of rationality that goes far beyond choice consistency, and has not received empirical support (Kurz 1997).

In real world applications of game theory, I conclude, we must have plausible grounds for believing that the equilibrium concept used is appropriate. Simply assuming that rationality implies Nash equilibrium, as is the case in classical game theory, is generally inappropriate. Evolutionary game theory restores the centrality of the Nash equilibrium concept, because stable equilibria of the replicator dynamic (and related "monotone" dynamics) are necessarily Nash equilibria. Moreover, the examples given in next section are restricted to games that are sufficiently simple that the sorts of anomalies discussed above are not present, and the Nash equilibrium criterion is appropriate.

10 Behavioral game theory and other-regarding preferences

Contemporary biological theory maintains that cooperation can be sustained by means of *inclusive fitness*, or cooperation among kin (Hamilton 1963), and by individual self-interest in the form of *reciprocal altruism* (Trivers 1971). Reciprocal altruism occurs when an individual helps another individual, at a fitness cost to itself, contingent on the beneficiary returning the favor in a future period. The explanatory power of inclusive fitness theory and reciprocal altruism convinced a generation of biologists that what appears to be altruism—personal sacrifice on behalf of others—is really just long-run genetic self-interest.[15] Combined with a vigorous critique of group selection (Williams 1966, Dawkins 1976, Maynard Smith 1976), a generation of biologists became convinced that true altruism—one organism sacrificing fitness

---

[14] One could posit that the "type" of a player must include the player's probability distribution over the types of other players, but even such arcane assumptions do not solve the problem.

[15] Current research is less sanguine concerning the important of reciprocal altruism in non-humans (Hammerstein 2003).

on behalf of the fitness of an unrelated other—was virtually unknown, even in the case of *Homo sapiens*.

That human nature is selfish was touted as a central implication of rigorous biological modeling. In *The Selfish Gene* (1976), for example, Richard Dawkins asserts that "We are survival machines—robot vehicles blindly programmed to preserve the selfish molecules known as genes.…Let us try to teach generosity and altruism, because we are born selfish." Similarly, in *The Biology of Moral Systems* (1987, p. 3), 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." More poetically, 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."

In economics, the notion that enlightened self-interest allows individuals to cooperate in large groups goes back to Bernard Mandeville's "private vices, public virtues" (1924[1705]) and Adam Smith's "invisible hand" (2000[1759]). Full analytical development of this idea awaited the Twentieth century development of general equilibrium theory (Arrow and Debreu 1954, Arrow and Hahn 1971) and the theory of repeated games (Axelrod and Hamilton 1981, Fudenberg and Maskin 1986).

By contrast, sociological, anthropological, and social psychological theory generally explain that human cooperation is predicated on affiliative behaviors among group members, each of whom is prepared to sacrifice a modicum of personal well-being to advance the group's collective goals. The vicious attack on "sociobiology" (Segerstrale 2001) and the widespread rejection of the bare-bones *Homo economicus* in the "soft" social sciences (Etzioni 1985, Hirsch, Michaels and Friedman 1990, DiMaggio 1994) is the result in part of this clash of basic explanatory principles.

Behavioral game theory assumes the BPC model, and it subjects individuals to strategic settings, such that their behavior reveals their underlying preferences. This controlled setting allows us to adjudicate between these contrasting models. One behavioral regularity that has been found thereby is *strong reciprocity*, which is a predisposition to cooperate with others, and to punish those who violate the norms of cooperation, at personal cost, even when it is implausible to expect that these costs will be repaid. Strong reciprocity is other-regarding, as a strong reciprocator's behavior reflects a preference to cooperate with other cooperators and to punish non-cooperators, even when these actions are personally costly.

The result of the laboratory and field research on strong reciprocity is that humans indeed often behave in ways that have traditionally been affirmed in sociological theory and denied in biology and economics (Ostrom, Walker and Gardner 1992, Andreoni 1995, Fehr, Gächter and Kirchsteiger 1997, Fehr, Kirchsteiger and Riedl 1998, Gächter and Fehr 1999, Fehr and Gächter 2000, Fehr and Gächter 2002, Henrich, Boyd, Bowles, Camerer, Fehr and Gintis 2005). Moreover, it is probable that this other-regarding behavior is a prerequisite for cooperation in large groups of non-kin, because the theoretical models of cooperation in large groups of self-regarding non-kin in biology and economics do not apply to some important and frequently observed forms of human cooperation (Boyd and Richerson 1992, Gintis 2005).

Another form of prosocial behavior conflicting with the maximization of personal material gain is that of maintaining such *character virtues* as honesty and promise-keeping, even when there is no chance of being penalized for unvirtuous behavior. An example of such behavior is reported by Gneezy (2005), who studied 450 undergraduate participants paired off to play three games of the following form. Player 1 would be shown two pairs of payoffs, A:(x, y) and B:(z, w) where x, y, z, and w are amounts of money with $x < z$ and $y > w$. Player 1 could then say to Player 2, who could not see the amounts of money, either “Option A will earn you more money than option B,”or “Option B will earn you more money than option A.” The first game was A:(5,6) vs. B:(6,5) so player 1 could gain 1 by lying and being believed, while imposing a cost of 1 on player 2. The second game was A:(5,15) vs. B:(6,5) so player 1 could gain 10 by lying and being believed, while still imposing a cost of 1 on player 2. The third game was A:(5,15) vs. B:(15,5), so player 1 could gain 10 by lying and being believed, while imposing a cost of 10 on player 2.

Before starting play, Gneezy asked Player 1’s whether they expected their advice to be followed, inducing honest responses by promising to reward subjects whose guesses were correct. He found that 82% of Player 1’s expected their advice to be followed (the actual number was 78%). It follows from the Player 1 expectations that if they were self-regarding, they would always lie and recommend B to Player 2. In fact, in game 2, where lying was very costly to Player 2 and the gain to lying for player 1 was small, only 17% of subjects lied. In game 1, where the cost of lying to Player 2 was only 1 but the gain to Player 1 was the same as in Game 2, 36% lied. In other words, subjects were loathe to lie, but considerably more so when it was costly to their partner. In game three, where the gain from lying was large for Player 1, and equal to the loss to Player 2, fully 52% lied. This shows that many subjects are willing to sacrifice material gain to avoid lying in a one-shot, anonymous interaction, their willingness to lie increasing with an increased cost of truth-telling to themselves, and decreasing with an increase in their partner’s cost of begin deceived. Similar results were found by Boles, Croson and Murnighan (2000) and Charness and Dufwenberg (2004). Gunnthorsdottir, McCabe and Smith (2002) and Burks, Carpenter and Verhoogen (2003) have shown that a social-psychological measure of “Machiavellianism” predicts which subjects are likely to be trustworthy and trusting.

11 Beliefs: the weak link in the BPC model

In the simplest formulation of the rational actor model, beliefs do not explicitly appear. In the real world, however, the probabilities of various outcomes in a lottery are rarely objectively known, and hence must generally be subjectively constructed as part of an individual's belief system. Anscombe and Aumann (1963) extended the Savage model to preferences over bundles consisting of "states of the world" and payoff bundles, and they showed that if certain consistency axioms hold, the individual could be modeled as maximizing subject to a set of subjective probabilities (beliefs) over states. Were these axioms universally plausible, beliefs could be derived in the same way as are preferences. However, at least one of these axioms, the so-called *state-independence axiom*, which states that preferences over payoffs are independent of the states in which they occur, is generally not plausible.

It follows that beliefs are the underdeveloped member of the BPC trilogy. Except for Bayes' rule (Gintis 2000a): Ch. 17, there is no compelling analytical theory of how a rational agent acquires and updates beliefs, although there are many partial theories (Kuhn 1962, Polya 1990, Boyer 2001, Jaynes 2003).

Beliefs enter the decision process in several potential ways. First, individuals may not have perfect knowledge concerning how their choices affect their welfare. This is most likely to be the case in an unfamiliar setting, of which the experimental laboratory is often a perfect example. In such cases, when forced to choose, individuals "construct" their preferences on the spot by forming beliefs based on whatever partial information is present at the time of choice (Slovic 1995). Understanding this process of belief formation is a demanding research task.

Second, often the actual actions $a \in A$ available to an individual will differ from the actual payoffs $\pi \in \Pi$ that appear in the individual's preference function. The mapping $\beta : A \to \Pi$ the individual deploys to maximize payoff is a belief system concerning objective reality, and it can differ from the correct mapping $\beta^* : A \to \Pi$. For example, a gambler may want to maximize expected winnings, but may believe in the erroneous Law of Small Numbers (Rabin 2002). Errors of this type include the *performance errors* discussed in section 9.6.

Third, there is considerable evidence that beliefs directly affect well-being, so individuals may alter their beliefs as part of their optimization program. Self-serving beliefs, unrealistic expectations, and projection of one's own preferences on others are important examples. The trade-off here is that erroneous beliefs may add to well-being, but acting on these beliefs may lower other payoffs (Bodner and Prelec 2002, Benabou and Tirole 2002).

12 Conclusion

Each of the behavioral disciplines contributes strongly to understanding human behavior. Taken separately and at face value, however, they offer partial, conflicting, and incompatible models. From a scientific point of view, it is scandalous that this situation was tolerated throughout most of the twentieth century. Fortunately, there is currently a strong current of unification based on both mathematical models and common methodological principles for gathering empirical data on human behavior and human nature.

The true power of each discipline's contribution to knowledge will only appear when suitably qualified and deepened by the contribution of the others. For example, the economist's model of rational choice behavior must be qualified by a biological appreciation that preference consistency is the result of strong evolutionary forces, and that where such forces are absent, consistency may be imperfect. Moreover, the notion that preferences are purely self-regarding must be abandoned. For a second example, the sociologist's notion of internalization of norms must be thoroughly integrated into behavioral theory, which must recognize that the ease with which diverse values can be internalized depends on human nature (Tooby and Cosmides 1992, Pinker 2002), and rate at which values are acquired and abandoned depends on their contribution to fitness and well-being (Gintis 2003b, Gintis 2003a), and there are often rapid society-wide value changes that cannot be accounted for by socialization theory at all (Wrong 1961, Gintis 1975).

Disciplinary boundaries in the behavioral sciences have been determined historically, rather than conforming to some consistent scientific logic. Perhaps for the first time, we are in a position to rectify this situation. We must recognize evolutionary theory (covering both genetic and cultural evolution) as the integrating principle of behavioral science. Moreover, if the BPC model is broadened to encompass other-regarding preferences, and a cogent theory of belief formation and change is developed, game theory becomes capable of modeling all aspects of decision making, including those normally considered "sociological" or "anthropological," which in turn is most naturally the central organizing principle of psychology.

I would like to thank George Ainslie, Rob Boyd, Dov Cohen, Ernst Fehr, Barbara Finlay, Thomas Getty, Dennis Krebs, Joe Henrich, Daniel Kahneman, Laurent Keller, Joachim Krueger, Larry Samuelson, and especially Marc Hauser and anonymous referees of this journal for helpful comments, and the John D. and Catherine T. MacArthur Foundation for financial support.

References

Abbott, R. J., J. K. James, R. I Milne, and A. C. M Gillies, “Plant Introductions, Hybridization and Gene Flow,” *Philosophical Transactions of the Royal Society of London B* 358 (2003):1123–1132.

Ahlbrecht, Martin and Martin Weber, “Hyperbolic Discounting Models in Prescriptive Theory of Intertemporal Choice,” *Zeitschrift für Wirtschafts- und Sozialwissenschaften* 115 (1995):535–568.

Ainslie, George, “Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control,” *Psychological Bulletin* 82 (July 1975):463–496.

— and Nick Haslam, “Hyperbolic Discounting,” in George Loewenstein and Jon Elster (eds.) *Choice Over Time* (New York: Russell Sage, 1992) pp. 57–92.

Akerlof, George A., “Procrastination and Obedience,” *American Economic Review* 81,2 (May 1991):1–19.

Alcock, John, *Animal Behavior: An Evolutionary Approach* (Sunderland, MA: Sinauer, 1993).

Alexander, R. D., *The Biology of Moral Systems* (New York: Aldine, 1987).

Allais, Maurice, “Le comportement de l’homme rationnel devant le risque, critique des postulats et axiomes de l’école Américaine,” *Econometrica* 21 (1953):503–546.

Allman, J., A. Hakeem, and K. Watson, “Two Phylogenetic Specializations in the Human Brain,” *Neuroscientist* 8 (2002):335–346.

Andreoni, James, “Cooperation in Public Goods Experiments: Kindness or Confusion,” *American Economic Review* 85,4 (1995):891–904.

— and John H. Miller, “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism,” *Econometrica* 70,2 (2002):737–753.

Anscombe, F. and R. Aumann, “A Definition of Subjective Probability,” *Annals of Mathematical Statistics* 34 (1963):199–205.

Arrow, Kenneth J. and Frank Hahn, *General Competitive Analysis* (San Francisco: Holden-Day, 1971).

— and Gerard Debreu, “Existence of an Equilibrium for a Competitive Economy,” *Econometrica* 22,3 (1954):265–290.

Atran, Scott, *In Gods We Trust* (Oxford: Oxford University Press, 2004).

Aumann, Robert and Adam Brandenburger, “Epistemic Conditions for Nash Equilibrium,” *Econometrica* 65,5 (September 1995):1161–80.

Axelrod, Robert and William D. Hamilton, “The Evolution of Cooperation,” *Science* 211 (1981):1390–1396.

Bandura, Albert, *Social Learning Theory* (Englewood Cliffs, NJ: Prentice Hall, 1977).

Becker, Gary S. and George J. Stigler, "De Gustibus Non Est Disputandum," *American Economic Review* 67,2 (March 1977):76–90.

— and Kevin M. Murphy, "A Theory of Rational Addiction," *Journal of Political Economy* 96,4 (August 1988):675–700.

— , Michael Grossman, and Kevin M. Murphy, "An Empirical Analysis of Cigarette Addiction," *American Economic Review* 84,3 (June 1994):396–418.

Beer, J. S., E. A. Heerey, D. Keltner, D. Skabini, and R. T. Knight, "The Regulatory Function of Self-conscious Emotion: Insights from Patients with Orbitofrontal Damage," *Journal of Personality and Social Psychology* 65 (2003):594–604.

Bell, D. E., "Regret in Decision Making under Uncertainty," *Operations Research* 30 (1982):961–981.

Benabou, Roland and Jean Tirole, "Self Confidence and Personal Motivation," *Quarterly Journal of Economics* 117,3 (2002):871–915.

Benedict, Ruth, *Patterns of Culture* (Boston: Houghton Mifflin, 1934).

Berg, Joyce E., John W. Dickhaut, and Thomas A. Rietz, "Preference Reversals: The Impact of Truth-Revealing Incentives," 2005. College of Business, University of Iowa.

Bernheim, B. Douglas, "Rationalizable Strategic Behavior," *Econometrica* 52,4 (July 1984):1007–1028.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welsh, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy* 100 (October 1992):992–1026.

Binmore, Ken, "Modelling Rational Players: I," *Economics and Philosophy* 3 (1987):179–214.

Black, Fisher and Myron Scholes, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81 (1973):637–654.

Blount, Sally, "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences," *Organizational Behavior & Human Decision Processes* 63,2 (August 1995):131–144.

Bodner, Ronit and Drazen Prelec, "Self-signaling and Diagnostic Utility in Everyday Decision Making," in Isabelle Brocas and Juan D. Carillo (eds.) *Collected Essays in Psychology and Economics* (Oxford: Oxford University Press, 2002) pp. 105–123.

Boehm, Christopher, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Cambridge, MA: Harvard University Press, 2000).

Boles, Terry L., Rachel T. A. Croson, and J. Keith Murnighan, "Deception and Retribution in Repeated Ultimatum Bargaining," *Organizational Behavior and Human Decision Processes* 83,2 (2000):235–259.

Bonner, John Tyler, *The Evolution of Culture in Animals* (Princeton, NJ: Princeton University Press, 1984).

Borgerhoff Mulder, Monique, "The Demographic Transition: Are we any Closer to an Evolutionary Explanation?," *Trends in Ecology and Evolution* 13,7 (July 1998):266–270.

Bowles, Samuel and Herbert Gintis, "Walrasian Economics in Retrospect," *Quarterly Journal of Economics (*November 2000):1411–1439.

— and — , "Prosocial Emotions," in Lawrence E. Blume and Steven N. Durlauf (eds.) *The Economy As an Evolving Complex System III* (Santa Fe, NM: Santa Fe Institute, 2005).

Boyd, Robert and Peter J. Richerson, *Culture and the Evolutionary Process* (Chicago: University of Chicago Press, 1985).

— and — , "The Evolution of Reciprocity in Sizable Groups," *Journal of Theoretical Biology* 132 (1988):337–356.

— and — , "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizeable Groups," *Ethology and Sociobiology* 113 (1992):171–195.

Boyer, Pascal, *Religion Explained: The Human Instincts That Fashion Gods, Spirits and Ancestors* (London: William Heinemann, 2001).

Brigden, Linda Waverley and Joy De Beyer, *Tobacco Control Policy: Stories from Around the World* (Washington, DC: World Bank, 2003).

Brown, Donald E., *Human Universals* (New York: McGraw-Hill, 1991).

Brown, J. H. and M. V. Lomolino, *Biogeography* (Sunderland, MA: Sinauer, 1998).

Burks, Stephen V., Jeffrey P. Carpenter, and Eric Verhoogen, "Playing Both Roles in the Trust Game," *Journal of Economic Behavior and Organization* 51 (2003):195–216.

Camerer, Colin, *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton, NJ: Princeton University Press, 2003).

Camille, N., "The Involvement of the Orbitofrontal Cortex in the Experience of Regret," *Science* 304 (2004):1167–1170.

Cavalli-Sforza, Luca L. and Marcus W. Feldman, "Theory and Observation in Cultural Transmission," *Science* 218 (1982):19–27.

Cavalli-Sforza, Luigi L. and Marcus W. Feldman, *Cultural Transmission and Evolution* (Princeton, NJ: Princeton University Press, 1981).

Charness, Gary and Martin Dufwenberg, "Promises and Partnership," October 2004. University of California and Santa Barbara.

Cheng, P. W. and K. J. Holyoak, “Pragmatic Reasoning Schemas,” *Cognitive Psychology* 17 (1985):391–416.

Cohen, L. Jonathan, “Can Human Irrationality be Experimentally Demonstrated?,” *Behavioral and Brain Sciences* 4 (1981):317–331.

Coleman, James S., *Foundations of Social Theory* (Cambridge, MA: Belknap, 1990).

Conlisk, John, “Optimization Cost,” *Journal of Economic Behavior and Organization* 9 (1988):213–228.

Cooper, W. S., “Decision Theory as a Branch of Evolutionary Theory,” *Psychological Review* 4 (1987):395–411.

Cosmides, Leda, “The Logic of Social Exchange: Has Natural Selection Shaped how Humans Reason? Studies with the Watson Selection Task,” *Cognition* 31 (1989):187–276.

Darwin, Charles, *The Origin of Species by Means of Natural Selection* (London: John Murray, 1872). 6th Edition.

Dawkins, Richard, *The Selfish Gene* (Oxford: Oxford University Press, 1976).

— , *The Extended Phenotype: The Gene as the Unit of Selection* (Oxford: Freeman, 1982).

DiMaggio, Paul, “Culture and Economy,” in Neil Smelser and Richard Swedberg (eds.) *The Handbook of Economic Sociology* (Princeton: Princeton University Press, 1994) pp. 27–57.

Dorris, Paul W. Glimcherand Michael C. and Hannah M. Bayer, “Physiological Utility Theory and the Neuroeconomics of Choice,” 2005. Center for Neural Science, New York University.

Durham, William H., *Coevolution: Genes, Culture, and Human Diversity* (Stanford: Stanford University Press, 1991).

Durkheim, Emile, *Suicide, a Study in Sociology* (New York: Free Press, 1951).

Ellsberg, Daniel, “Risk, Ambiguity, and the Savage Axioms,” *Quarterly Journal of Economics* 75 (1961):643–649.

Elster, Jon, *Ulysses and the Sirens: Studies in Rationality and Irrationality* (Cambridge, UK: Cambridge University Press, 1979).

Eshel, Ilan and Marcus W. Feldman, “Initial Increase of New Mutants and Some Continuity Properties of ESS in two Locus Systems,” *American Naturalist* 124 (1984):631–640.

— , — , and Aviv Bergman, “Long-term Evolution, Short-term Evolution, and Population Genetic Theory,” *Journal of Theoretical Biology* 191 (1998):391–396.

Etzioni, Amitai, “Opening the Preferences: A Socio-Economic Research Agenda,” *Journal of Behavioral Economics* 14 (1985):183–205.

Fehr, Ernst and Simon Gächter, "Cooperation and Punishment," *American Economic Review* 90,4 (September 2000):980–994.

— and — , "Altruistic Punishment in Humans," *Nature* 415 (10 January 2002):137–140.

— , Georg Kirchsteiger, and Arno Riedl, "Gift Exchange and Reciprocity in Competitive Experimental Markets," *European Economic Review* 42,1 (1998):1–34.

— , Simon Gächter, and Georg Kirchsteiger, "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica* 65,4 (July 1997):833–860.

Feldman, Marcus W. and Lev A. Zhivotovsky, "Gene-Culture Coevolution: Toward a General Theory of Vertical Transmission," *Proceedings of the National Academy of Sciences* 89 (December 1992):11935–11938.

Fishburn, Peter C. and Ariel Rubinstein, "Time Preference," *Econometrica* 23,3 (October 1982):667–694.

Fisher, Ronald A., *The Genetical Theory of Natural Selection* (Oxford: Clarendon Press, 1930).

Fudenberg, Drew and Eric Maskin, "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica* 54,3 (May 1986):533–554.

— , David K. Levine, and Eric Maskin, "The Folk Theorem with Imperfect Public Information," *Econometrica* 62 (1994):997–1039.

Gächter, Simon and Ernst Fehr, "Collective Action as a Social Exchange," *Journal of Economic Behavior and Organization* 39,4 (July 1999):341–369.

Gadagkar, Raghavendra, "On Testing the Role of Genetic Asymmetries Created by Haplodiploidy in the Evolution of Eusociality in the Hymenoptera," *Journal of Genetics* 70,1 (April 1991):1–31.

Ghiselin, Michael T., *The Economy of Nature and the Evolution of Sex* (Berkeley: University of California Press, 1974).

Gigerenzer, Gerd and Reinhard Selten, *Bounded Rationality* (Cambridge, MA: MIT Press, 2001).

Gilovich, T., R. Vallone, and A. Tversky, "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Journal of Personality and Social Psychology* 17 (1985):295–314.

Gintis, Herbert, "A Radical Analysis of Welfare Economics and Individual Development," *Quarterly Journal of Economics* 86,4 (November 1972):572–599.

— , "Welfare Economics and Individual Development: A Reply to Talcott Parsons," *Quarterly Journal of Economics* 89,2 (February 1975):291–302.

— , *Game Theory Evolving* (Princeton, NJ: Princeton University Press, 2000).

— , “Strong Reciprocity and Human Sociality,” *Journal of Theoretical Biology* 206 (2000):169–179.

— , “The Hitchhiker’s Guide to Altruism: Genes, Culture, and the Internalization of Norms,” *Journal of Theoretical Biology* 220,4 (2003):407–418.

— , “Solving the Puzzle of Human Prosociality,” *Rationality and Society* 15,2 (May 2003):155–187.

— , “The Competitive Economy as a Complex Dynamical System,” 2004. Santa Fe Institute Working Paper.

— , “Behavioral Game Theory and Contemporary Economic Theory,” *Analyze & Kritik* 27,1 (2005):48–72.

— , “The Evolution of Private Property,” *Journal of Economic Behavior and Organization (*2006).

— , Samuel Bowles, Robert Boyd, and Ernst Fehr, *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life* (Cambridge: The MIT Press, 2005).

Glimcher, Paul W., *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics* (Cambridge, MA: MIT Press, 2003).

Gneezy, Uri, “Deception: The Role of Consequences,” *American Economic Review* 95,1 (March 2005):384–394.

Goldstein, E. Bruce, *Cognitive Psychology: Connecting Mind, Research, and Everyday Experience* (New York: Wadsworth, 2005).

Grafen, Alan, “Formal Darwinism, the Individual-as-maximizing-agent Analogy, and Bet-hedging,” *Proceedings of the Royal Society B* 266 (1999):799–803.

— , “Developments of Price’s Equation and Natural Selection Under Uncertainty,” *Proceedings of the Royal Society B* 267 (2000):1223–1227.

— , “A First Formal Link between the Price Equation and an Optimization Program,” *Journal of Theoretical Biology* 217 (2002):75–91.

Grether, David and Charles Plott, “Economic Theory of Choice and the Preference Reversal Phenomenon,” *American Economic Review* 69,4 (September 1979):623–638.

Grice, H. P., “Logic and Conversation,” in Donald Davidson and Gilbert Harman (eds.) *The Logic of Grammar* (Encino, CA: Dickenson, 1975) pp. 64–75.

Gruber, J. and B. Koszegi, “Is Addiction Rational? Theory and Evidence,” *Quarterly Journal of Economics* 116,4 (2001):1261–1305.

Grusec, Joan E. and Leon Kuczynski, *Parenting and Children’s Internalization of Values: A Handbook of Contemporary Theory* (New York: John Wily & Sons, 1997).

Gunnthorsdottir, Anna, Kevin McCabe, and Vernon Smith, "Using the Machiavellianism Instrument to Predict Trustworthiness in a Bargaining Game," *Journal of Economic Psychology* 23 (2002):49–66.

Haldane, J. B. S., *The Causes of Evolution* (London: Longmans, Green & Co., 1932).

Hamilton, William D., "The Evolution of Altruistic Behavior," *American Naturalist* 96 (1963):354–356.

Hammerstein, Peter, "Darwinian Adaptation, Population Genetics and the Streetcar Theory of Evolution," *Journal of Mathematical Biology* 34 (1996):511–532.

— , "Why Is Reciprocity So Rare in Social Animals?," in Peter Hammerstein (ed.) *Genetic and Cultural Evolution of Cooperation* (Cambridge, MA: The MIT Press, 2003) pp. 83–93.

— and Reinhard Selten, "Game Theory and Evolutionary Biology," in Robert J. Aumann and Sergiu Hart (eds.) *Handbook of Game Theory with Economic Applications* (Amsterdam: Elsevier, 1994) pp. 929–993.

Harsanyi, John C., "Games with Incomplete Information Played by Bayesian Players, Parts I, II, and III," *Behavioral Science* 14 (1967):159–182, 320–334, 486–502.

Hechter, Michael and Satoshi Kanazawa, "Sociological Rational Choice," *Annual Review of Sociology* 23 (1997):199–214.

Heiner, Ronald A., "The Origin of Predictable Behavior," *American Economic Review* 73,4 (1983):560–595.

Henrich, Joe, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis, "Economic Man' in Cross-Cultural Perspective: Behavioral Experiments in 15 small-scale societies," *Behavioral and Brain Sciences (*2005).

Henrich, Joseph, "Market Incorporation, Agricultural Change and Sustainability among the Machiguenga Indians of the Peruvian Amazon," *Human Ecology* 25,2 (June 1997):319–351.

— , "Cultural Transmission and the Diffusion of Innovations," *American Anthropologist* 103 (2001):992–1013.

— and Francisco Gil-White, "The Evolution of Prestige: Freely Conferred Status as a Mechanism for Enhancing the Benefits of Cultural Transmission," *Evolution and Human Behavior* 22 (2001):1–32.

— and Robert Boyd, "The Evolution of Conformist Transmission and the Emergence of Between-Group Differences," *Evolution and Human Behavior* 19 (1998):215–242.

Herrnstein, Richard, David Laibson, and Howard Rachlin, *The Matching Law: Papers on Psychology and Economics* (Cambridge, MA: Harvard University

Press, 1997).

Herrnstein, Richard J., "Relative and Absolute Strengths of Responses as a Function of Frequency of Reinforcement," *Journal of Experimental Analysis of Animal Behavior* 4 (1961):267–272.

Hilton, Denis J., "The Social Context of Reasoning: Conversational Inference and Rational Judgment," *Psychological Bulletin* 118,2 (1995):248–271.

Hirsch, Paul, Stuart Michaels, and Ray Friedman, "Clean Models vs. Dirty Hands: Why Economics is Different from Sociology," in Sharon Zukin and Paul DiMaggio (eds.) *Structures of Capital: The Social Organization of the Economy* (New York: Cambridge University Press, 1990) pp. 39–56.

Holden, C. J., "Bantu Language Trees Reflect the Spread of Farming Across Sub-Saharan Africa: A Maximum-parsimony Analysis," *Proceedings of the Royal Society of London Series B* 269 (2002):793–799.

— and Ruth Mace, "Spread of Cattle Led to the Loss of Matrilineal Descent in Africa: A Coevolutionary Analysis," *Proceedings of the Royal Society of London Series B* 270 (2003):2425–2433.

Huang, Chi-Fu and Robert H. Litzenberger, *Foundations for Financial Economics* (Amsterdam: Elsevier, 1988).

Huxley, Julian S., "Evolution, Cultural and Biological," *Yearbook of Anthropology (*1955):2–25.

Jablonka, Eva and Marion J. Lamb, *Epigenetic Inheritance and Evolution: The Lamarckian Case* (Oxford: Oxford University Press, 1995).

James, William, "Great Men, Great Thoughts, and the Environment," *Atlantic Monthly* 46 (1880):441–459.

Jaynes, E. T., *Probability Theory: The Logic of Science* (Cambridge: Cambridge University Press, 2003).

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica* 47 (1979):263–291.

— and — , *Choices, Values, and Frames* (Cambridge: Cambridge University Press, 2000).

— , Paul Slovic, and Amos Tversky, *Judgment under Uncertainty: Heuristics and Biases* (Cambridge, UK: Cambridge University Press, 1982).

Kiyonari, Toko, Shigehito Tanida, and Toshio Yamagishi, "Social Exchange and Reciprocity: Confusion or a Heuristic?," *Evolution and Human Behavior* 21 (2000):411–427.

Kollock, Peter, "Transforming Social Dilemmas: Group Identity and Cooperation," in Peter Danielson (ed.) *Modeling Rational and Moral Agents* (Oxford: Oxford University Press, 1997).

Krantz, D. H., "From Indices to Mappings: The Representational Approach to Measurement," in D. Brown and J. Smith (eds.) *Frontiers of Mathematical Psychology* (Cambridge: Cambridge University Press, 1991) pp. 1–52.

Krebs, J. R. and N. B. Davies, *Behavioral Ecology: An Evolutionary Approach* fourth ed. (Oxford: Blackwell Science, 1997).

Kreps, David M., *A Course in Microeconomic Theory* (Princeton, NJ: Princeton University Press, 1990).

Krueger, Joachim I. and David C. Funder, "Towards a balanced social psychology: Causes, Consequences, and Cures for the Problem-seeking Approach to Social Behavior and Cognition," *Behavioral and Brain Sciences* 27,3 (June 2004):313–327.

Kuhn, Thomas, *The Nature of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).

Kurz, Mordecai, "Endogenous Economic Fluctuations and Rational Beliefs: A General Perspective," in Mordecai Kurz (ed.) *Endogenous Economic Fluctuations: Studies in the Theory of Rational Beliefs* (Berlin: Springer-Verlag, 1997) pp. 1–37.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics* 112,2 (May 1997):443–477.

— , James Choi, and Brigitte Madrian, "Plan Design and 401(k) Savings Outcomes," *National Tax Journal* 57 (June 2004):275–298.

Lewontin, Richard C., *The Genetic Basis of Evolutionary Change* (New York: Columbia University Press, 1974).

Liberman, Uri, "External Stability and ESS Criteria for Initial Increase of a New Mutant Allele," *Journal of Mathematical Biology* 26 (1988):477–485.

Lichtenstein, Sarah and Paul Slovic, "Reversals of Preferences Between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology* 89 (1971):46–55.

Loomes, G. and Robert Sugden, "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty," *Economic Journal* 92 (1982):805–824.

Lumsden, C. J. and E. O. Wilson, *Genes, Mind, and Culture: The Coevolutionary Process* (Cambridge, MA: Harvard University Press, 1981).

Mace, Ruth and Mark Pagel, "The Comparative Method in Anthropology," *Current Anthropology* 35 (1994):549–564.

Mandeville, Bernard, *The Fable of the Bees: Private Vices, Publick Benefits* (Oxford: Clarendon, 1924[1705]).

Maynard Smith, John, "Group Selection," *Quarterly Review of Biology* 51 (1976):277–283.

— , *Evolution and the Theory of Games* (Cambridge, UK: Cambridge University Press, 1982).

— and Eors Szathmary, *The Major Transitions in Evolution* (Oxford: Oxford University Press, 1997).

Mazur, James E., *Learning and Behavior* (Upper Saddle River, NJ: Prentice-Hall, 2002).

McClure, Samuel M., David I. Laibson, George Loewenstein, and Jonathan D. Cohen, "Separate Neural Systems Value Immediate and Delayed Monetary Rewards," *Science* 306,5695 (15 October 2004):503–507.

McKelvey, R. D. and T. R. Palfrey, "An Experimental Study of the Centipede Game," *Econometrica* 60 (1992):803–836.

Mead, Margaret, *Sex and Temperament in Three Primitive Societies* (New York: Morrow, 1963).

Meltzhoff, Andrew N. and J. Decety, "What Imitation Tells us About Social Cognition: A Rapprochement Between Developmental Psychology and Cognitive Neuroscience," *Philosophical Transactions of the Royal Society of London B* 358 (2003):491–500.

Mesoudi, Alex, Andrew Whiten, and Kevin N. Laland, "Towards a Unified Science of Cultural Evolution," *Behavioral and Brain Sciences (*2006).

Miller, B. L., A. Darby, D. F. Benson, J. L. Cummings, and M. H. Miller, "Aggressive, Socially Disruptive and Antisocial Behaviour Associated with Frontotemporal Dementia," *British Journal of Psychiatry* 170 (1997):150–154.

Mischel, W., "Process in Delay of Gratification," in L. Berkowitz (ed.) *Advances in Experimental Social Psychology, Volume 7* (New York: Academic Press, 1974).

Moll, Jorge, Roland Zahn, Ricardo di Oliveira-Souza, Frank Krueger, and Jordan Grafman, "The Neural Basis of Human Moral Cognition," *Nature Neuroscience* 6 (October 2005):799–809.

Montague, P. Read and Gregory S. Berns, "Neural Economics and the Biological Substrates of Valuation," *Neuron* 36 (2002):265–284.

Moore, Jr., Barrington, *Injustice: The Social Bases of Obedience and Revolt* (White Plains: M. E. Sharpe, 1978).

Moran, P. A. P., "On the Nonexistence of Adaptive Topographies," *Annals of Human Genetics* 27 (1964):338–343.

Newman, Mark, Albert-Laszlo Barabasi, and Duncan J. Watts, *The Structure and Dynamics of Networks* (Princeton, NJ: Princeton University Press, 2006).

Nisbett, Richard E. and Dov Cohen, *Culture of Honor: The Psychology of Violence in the South* (Boulder: Westview Press, 1996).

O'Brian, M. J. and R. L. Lyman, *Applying Evolutionary Archaeology* (New York: Kluwer Academic, 2000).

Odling-Smee, F. John, Keven N. Laland, and Marcus W. Feldman, *Niche Construction: The Neglected Process in Evolution* (Princeton: Princeton University Press, 2003).

O'Donoghue, Ted and Matthew Rabin, "Choice and Procrastination," *Quarterly Journal of Economics* 116,1 (February 2001):121–160.

Olson, Mancur, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, MA: Harvard University Press, 1965).

Ostrom, Elinor, James Walker, and Roy Gardner, "Covenants with and without a Sword: Self-Governance Is Possible," *American Political Science Review* 86,2 (June 1992):404–417.

Parker, A. J. and W. T. Newsome, "Sense and the Single Neuron: Probing the Physiology of Perception," *Annual Review of Neuroscience* 21 (1998):227–277.

Parsons, Talcott, "Evolutionary Universals in Society," *American Sociological Review* 29,3 (June 1964):339–357.

— , *Sociological Theory and Modern Society* (New York: Free Press, 1967).

— and Edward Shils, *Toward a General Theory of Action* (Cambridge, MA: Harvard University Press, 1951).

Pearce, David, "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica* 52 (1984):1029–1050.

Pinker, Steven, *The Blank Slate: The Modern Denial of Human Nature* (New York: Viking, 2002).

Plott, Charles R., "The Application of Laboratory Experimental Methods to Public Choice," in Clifford S. Russell (ed.) *Collective Decision Making: Applications from Public Choice Theory* (Baltimore, MD: Johns Hopkins University Press, 1979) pp. 137–160.

Polya, George, *Patterns of Plausible Reasoning* (Princeton: Princeton University Press, 1990).

Popper, Karl, *Objective knowledge: An Evolutionary Approach* (Oxforc: Clarendon Press, 1979).

Poundstone, William, *Prisoner's Dilemma* (New York: Doubleday, 1992).

Power, T. G. and M. L. Chapieski, "Childrearing and Impulse Control in Toddlers: A Naturalistic Investigation," *Developmental Psychology* 22 (1986):271–275.

Rabin, Matthew, "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics* 117,3 (August 2002):775–816.

Real, Leslie A., "Animal Choice Behavior and the Evolution of Cognitive Architecture," *Science* 253 (30 August 1991):980–986.

Real, Leslie and Thomas Caraco, "Risk and Foraging in Stochastic Environments," *Annual Review of Ecology and Systematics* 17 (1986):371–390.

Richerson, Peter J. and Robert Boyd, "The Evolution of Ultrasociality," in I. Eibl-Eibesfeldt and F.K. Salter (eds.) *Indoctrinability, Idology and Warfare* (New York: Berghahn Books, 1998) pp. 71–96.

— and — , *Not By Genes Alone* (Chicago: University of Chicago Press, 2004).

Rivera, M. C. and J. A. Lake, "The Ring of Life Provides Evidence for a Genome Fusion Origin of Eukaryotes," *Nature* 431 (2004):152–155.

Rizzolatti, G., L. Fadiga, L Fogassi, and V. Gallese, "From Mirror Neurons to Imitation: Facts and Speculations," in Andrew N. Meltzhoff and Wolfgang Prinz (eds.) *The Imitative Mind: Development, Evolution and Brain Bases* (Cambridge: Cambridge University Press, 2002) pp. 247–266.

Rogers, Alan, "Evolution of Time Preference by Natural Selection," *American Economic Review* 84,3 (June 1994):460–481.

Rosenthal, Robert W., "Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox," *Journal of Economic Theory* 25 (1981):92–100.

Rozin, Paul, L. Lowery, S. Imada, and Jonathan Haidt, "The CAD Triad Hypothesis: A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity)," *Journal of Personality & Social Psychology* 76 (1999):574–586.

Saffer, Henry and Frank Chaloupka, "The Demand for Illicit Drugs," *Economic Inquiry* 37,3 (1999):401–11.

Sally, David, "Conversation and Cooperation in Social Dilemmas," *Rationality and Society* 7,1 (January 1995):58–92.

Schall, J. D. and K. G. Thompson, "Neural Selection and Control of Visually Guided Eye Movements," *Annual Review of Neuroscience* 22 (1999):241–259.

Schrödinger, Edwin, *What is Life?: The Physical Aspect of the Living Cell* (Cambridge: Cambridge University Press, 1944).

Schulkin, J., *Roots of Social Sensitivity and Neural Function* (Cambridge, MA: MIT Press, 2000).

Schultz, W., P. Dayan, and P. R. Montague, "A Neural Substrate of Prediction and Reward," *Science* 275 (1997):1593–1599.

Seeley, Thomas D., "Honey Bee Colonies are Group-Level Adaptive Units," *The American Naturalist* 150 (1997):S22–S41.

Segerstrale, Ullica, *Defenders of the Truth: The Sociobiology Debate* (Oxford: Oxford University Press, 2001).

Selten, Reinhard, "In Search of a Better Understanding of Economic Behavior," in Arnold Heertje (ed.) *The Makers of Modern Economics,* vol. 1 (Harvester

Wheatsheaf, 1993) pp. 115–139.

Shafir, Eldar and Robyn A. LeBoeuf, "Rationality," *Annual Review of Psychology* 53 (2002):491–517.

Shennan, Stephen, *Quantifying Archaeology* (Edinburgh: Edinburgh University Press, 1997).

Simon, Herbert, "Theories of Bounded Rationality," in C. B. McGuire and Roy Radner (eds.) *Decision and Organization* (New York: American Elsevier, 1972) pp. 161–176.

— , *Models of Bounded Rationality* (Cambridge, MA: MIT Press, 1982).

Skibo, James M. and R. Alexander Bentley, *Complex Systems and Archaeology* (Salt Lake City: University of Utah Press, 2003).

Slovic, Paul, "The Construction of Preference," *American Psychologist* 50,5 (1995):364–371.

Smith, Adam, *The Theory of Moral Sentiments* (New York: Prometheus, 2000[1759]).

Smith, Eric Alden and B. Winterhalder, *Evolutionary Ecology and Human Behavior* (New York: Aldine de Gruyter, 1992).

Smith, Vernon, "Microeconomic Systems as an Experimental Science," *American Economic Review* 72 (December 1982):923–955.

Stanovich, Keith E., *Who is Rational? Studies in Individual Differences in Reasoning* (New York: Lawrence Erlbaum Associates, 1999).

Stephens, W., C. M. McLinn, and J. R. Stevens, "Discounting and Reciprocity in an Iterated Prisoner's Dilemma," *Science* 298 (13 December 2002):2216–2218.

Sternberg, Robert J. and Richard K. Wagner, *Readings in Cognitive Psychology* (Belmont, CA: Wadsworth, 1999).

Sugden, Robert, "An Axiomatic Foundation for Regret Theory," *Journal of Economic Theory* 60,1 (June 1993):159–180.

Sugrue, Leo P., Gregory S. Corrado, and William T. Newsome, "Choosing the Greater of Two Goods: Neural Currencies for Valuation and Decision Making," *Nature Reviews Neuroscience* 6 (2005):363–375.

Sutton, R. and A. G. Barto, *Reinforcement Learning* (Cambridge, MA: The MIT Press, 2000).

Taylor, P. and L. Jonker, "Evolutionarily Stable Strategies and Game Dynamics," *Mathematical Biosciences* 40 (1978):145–156.

Tomasello, Michael, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll, "Understanding and Sharing Intentions: The Origins of Cultural Cognition," *Behavioral and Brain Sciences* 28,5 (2005):675–691.

Tooby, John and Leda Cosmides, "The Psychological Foundations of Culture," in Jerome H. Barkow, Leda Cosmides, and John Tooby (eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (New York: Oxford University Press, 1992) pp. 19–136.

Trivers, Robert L., "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology* 46 (1971):35–57.

Tversky, Amos and Daniel Kahneman, "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics* 106,4 (November 1981):1039–1061.

— and Daniel Kanheman, "Belief in the Law of Small Numbers," *Psychological Bulletin* 76 (1971):105–110.

— and — , "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgement," *Psychological Review* 90 (1983):293–315.

— , Paul Slovic, and Daniel Kahneman, "The Causes of Preference Reversal," *American Economic Review* 80,1 (March 1990):204–217.

Von Neumann, John and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press, 1944).

Wason, P. C., "Reasoning," in B. Foss (ed.) *New Horizons in Psychology* (Harmonsworth: Penguin, 1966) pp. 135–151.

Wetherick, N. E., "Reasoning and Rationality: A Critique of Some Experimental Paradigms," *Theory & Psychology* 5,3 (1995):429–448.

Williams, G. C., *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton, NJ: Princeton University Press, 1966).

Williams, J. H. G., A. Whiten, T. Suddendorf, and D. I Perrett, "Imitation, Mirror Neurons and Autism," *Neuroscience and Biobheavioral Reviews* 25 (2001):287–295.

Wilson, Edward O., *Consilience: The Unity of Knowledge* (New York: Knopf, 1998).

Winter, Sidney G., "Satisficing, Selection and the Innovating Remnant," *Quarterly Journal of Economics* 85 (1971):237–261.

Wood, Elisabeth Jean, *Insurgent Collective Action and Civil War in El Salvador* (Cambridge,: Cambridge University Press, 2003).

Wright, Sewall, "Evolution in Mendelian Populations," *Genetics* 6 (1931):111–178.

Wrong, Dennis H., "The Oversocialized Conception of Man in Modern Sociology," *American Sociological Review* 26 (April 1961):183–193.

Young, H. Peyton, *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions* (Princeton, NJ: Princeton University Press, 1998).

Zajonc, R. B., “Feeling and Thinking: Preferences Need No Inferences,” *American Psychologist* 35,2 (1980):151–175.

Zajonc, Robert B., “On the Primacy of Affect,” *American Psychologist* 39 (1984):117–123.

c\Papers\Unity of Science\UnityBBS.tex May 6, 2006