BEHAVIORAL AND BRAIN SCIENCES (1991) 14, 205–245
Printed in the United States of America

The quest for optimality: A positive heuristic of science?

Paul J. H. Schoemaker
Center for Decision Research, Graduate School of Business, University of Chicago, Chicago, IL 60637
Electronic mail: *fac_paul@gsbacd.uchicago.edu*

Abstract: This paper examines the strengths and weaknesses of one of science's most pervasive and flexible metaprinciples; *optimality* is used to explain utility maximization in economics, least effort principles in physics, entropy in chemistry, and survival of the fittest in biology. Fermat's principle of least time involves both teleological and causal considerations, two distinct modes of explanation resting on poorly understood psychological primitives. The rationality heuristic in economics provides an example from social science of the potential biases arising from the extreme flexibility of optimality considerations, including selective search for confirming evidence, ex post rationalization, and the confusion of prediction with explanation. Commentators are asked to reflect on the extent to which optimality is (1) an organizing principle of nature, (2) a set of relatively unconnected techniques of science, (3) a normative principle for rational choice and social organization, (4) a metaphysical way of looking at the world, or (5) something else still.

Keywords: adaptation; biases; causality; control theory; economics; entropy; evolution; explanation; heuristics; homeostasis; optimization; rationality; regulation; sociobiology; variational principles

Most scientists seek to *explain* as well as describe and predict empirical regularities.[1] Often such explanations involve *optimality arguments*, ranging from "least effort" principles in physics to "survival of the fittest" in biology or economics. This paper suggests that the use of optimality in science is a powerful heuristic (i.e., a mental shortcut) for describing existing phenomena as well as predicting new ones. As with any heuristic (Tversky & Kahneman 1974), however, the optimality approach is prone to systematic biases, which the paper identifies.

First, a brief review is offered of the pervasiveness of optimality arguments in various sciences. The concept of optimality is then examined as a model bound notion, involving intentional and teleological perspectives. Fermat's principle of least time is used to examine the difference between teleological and causal theories, followed by a discussion of modes of scientific explanation. The paper closes with an examination of the rationality assumption underlying economics. Both the power of the optimality heuristic and its inherent biases are highlighted.

1. Optimality principles

Whenever a behavior or other empirical phenomenon is explained as maximizing or minimizing some objective function (subject to well-defined constraints), an optimality principle is implicitly or explicitly adduced. This paper examines the scientific use of such optimality principles, a key characteristic of which is that the empirical phenomenon of interest is viewed as a necessary consequence of optimizing some well-specified objective function. The following are examples of such optimality principles in various fields of inquiry. To develop the argument, we will equate extremum principles with optimality principles (although some readers may find this objectionable).

1.1. Economics. Among the social sciences, economics is most closely wedded to the optimality approach, especially at the microlevel. Individual consumers as well as business organizations are presumed to be maximizing entities who calculate with lightning speed their optimal consumption patterns and output levels. Cournot (1838), Pareto (1897), and others (such as Edgeworth, Slutsky, Walras, and Marshall) introduced mathematics with great vigor (and controversy) into economics, and Paul Samuelson (1946) set the standard for subsequent generations regarding the use of formal analysis. Today optimality models abound in economics, ranging from Pareto optimality to the optimal designs of incentive structures and contracts in firms.

Current theories of finance, as a branch of microeconomics, offer a good example of the approach (Fama 1976). Firms are presumed to issue stock and debt in ratios that minimize the total cost of capital. Investors assess stock prices by rationally projecting dividends and discounting them for time and risk. The latter concerns only the so-called systematic risk component (i.e., covariance with the market) as most firm-specific risk can be diversified away via portfolios. Investors are assumed to hold only *efficient* portfolios, that is, those having minimum risk for a given level of expected return. New information immediately gives rise to new expectations concerning dividends (via Bayes' theorem), so that stock prices follow Martingale distributions or random walks (see Fama & Miller 1972). Although no claim is made that

 0140-525X/91 $5.00+.00

anyone actually solves the complex equations involved, it is nonetheless argued (in the tradition of positivism) that such "as if" assumptions closely predict aggregate real-world behavior.

Several authors have examined optimality parallels between economics and physics (Magill 1970; Samuelson 1970; Tinbergen 1928) as well as economics and biology (Alchian 1950; Cooper 1987; Ghiselin 1974; Hirshleifer 1977; Houthakker 1956; Maynard Smith 1978). In economics, either expected utility or economic profit is maximized; in biology, mean fitness or reproductive survival of genes or organisms (Lewontin 1974).[2] In the latter case (as in physics and chemistry), the optimization argument is clearly advanced "as if," since biological entities lack the conscious striving and foresight characteristic of humans. Thus, the justification for the optimality heuristic seems, a fortiori, stronger in economics than in the life or physical sciences, although some consider even economic rationality no more than "as if" (Friedman 1953).

1.2. Physics. Maupertuis's (1744) principle of least action is perhaps the first major use of a formal optimality argument in science (see von Helmholtz 1886). It holds that a mechanical system moves along the path of least resistance (i.e., minimal mechanical action). In optics, this law was known earlier (ca. 1657) as Fermat's principle of least time (assuming total energy remains constant). Both principles were later generalized, following Euler (1744) and Lagrange (1788), by Hamilton (1834; 1835) to systems without conservation of energy. Hamilton's principle of least action has become one of the major unifying concepts of theoretical physics.

The application of Hamilton's principle involves the calculus of variations and can be illustrated by considering the trajectory a ball will follow when thrown away in the air. In a constant gravitational field, without other forces operating, the trajectory of the ball will be a parabola. This is derivable from the least action principle as follows. Let $\frac{1}{2}mv(x)^2$ be the kinetic energy at a given point x along the trajectory and mgx its potential energy (where m is mass, v velocity and g the gravitational constant). If we "sum" the differences between kinetic and potential energy along the trajectory, the following function obtains:

$$\text{Action} = \int_{t_1}^{t_2} [\frac{1}{2}mv(t)^2 - mgx(t)]dt \quad \text{where } x = f(t) \text{ and } v = \frac{dx}{dt}$$

To find the path of least action, perturbation analysis can be used along a presumed optimal path $x(t)$. This method, familiar in the calculus of variations, will yield a parabolic function as the solution (see Hylleraas 1970).

Similar calculus of variation models are found in relativity theory, where the optimal path (in curved space) corresponds to that of minimum length. In membrane physics, the minimum energy principle translates into liquid films enveloping a given volume (e.g., droplet) with minimum surface area. In electrostatics it predicts that the potential between two conductors adjusts itself so that electrostatic energy is minimal. Similarly, an electric current will distribute itself through a material so as to minimize heat generation (assuming Ohm's law holds). As Feynman et al. (1964, Chapter 19, p. 13) noted, Hamilton's law is a cornerstone principle of both classical and modern physics, which offers "excellent numerical results for otherwise intractable problems."

1.3. Chemistry. Apart from its use of physics, chemistry has developed optimality principles of its own (although fewer than physics). Perhaps the most general one is the equilibrium concept in chemical kinetics. Equilibrium seeking can be viewed and expressed as minimizing a difference function defined on the actual and ideal states. Le Chattelier's principle, for example, predicts that solutions in equilibrium will react chemically to counter the cause of any change that might be introduced (see Brescia et al. 1966, pp. 341–43). Another example in chemistry/physics is Hund's rule of maximum multiplicity, which describes how electrons fill up orbits around atoms (Brescia et al. 1966, p. 189).

A second major principle in chemistry (and physics) is that of entropy maximization in closed systems. When two gases are mixed, this principle predicts that their configuration will tend toward maximum chaos. The entropy concept is closely linked to that of maximum likelihood, which features prominently in statistical mechanics and physical chemistry (e.g., the Maxwell-Boltzmann distribution law for molecular velocities in a gas; see Tipler 1969, pp. 66–78). In contrast to the least effort principle, chemistry advanced in the late nineteenth century the Thomsen-Berthelot principle of *maximum* work or heat to explain chemical reactions between solids (see Partington 1964). Later, this principle was restated in terms of free energy and entropy.

1.4. Biology. Although biology seems qualitatively different from physics and chemistry in that it examines mutable living systems, many links exist. The equilibrium-seeking laws of chemistry and physics find expression in dissipative structures as the principle of homeostasis. The latter causes the cells or organism to maintain certain chemical balances and tree leaves to position themselves for optimal sun intake. Homeostasis may also underlie various allometric laws, which describe optimal relationships between form and function (d'Arcy Thompson 1917; Varela 1979). For example, stable log-linear relationships can be found between head (or tail) length of fish and total length; or when plotting trunk width against trunk length in adult mammals. The presumption is that these ratios reflect optimal design and biological autonomy.[3]

The presumed driving force behind such biological optimality is natural selection. Whenever a population possesses (1) variance in genotype, (2) inheritability of genotype, and (3) natural selection, evolution is predicted toward the fittest (in the reproductive sense). Much evidence exists for such evolution, which in stable environments might bring forth optimal adaptation. For example, the shark, barracuda, and dolphin exhibit remarkable similarity of form (for such different vertebrates), suggesting that their common environment has evolved an "optimal" design. Similarly, it has been mathematically argued that our vascular system has evolved into an optimal network when examining at what angles arteries branch, how often they branch, and how diameters narrow further down the supply lines (Rosen 1967).

Ecology or population biology is especially drawn to the optimality approach. Various analytic and empirical studies have claimed optimal sex ratios in species (Bull 1983; Karlin & Lessard 1986), optimal foraging (Charnov 1976; Fantino & Abarca 1985; Rodman & Cant 1984) and predator switching (Rapport 1971) in animals, optimal division of labor among social insects (Wilson 1975), or optimal mating behavior in animals (Maynard Smith 1982). Strong parallels exist, in these models, with those of economics where selection occurs at the organizational as well as individual level. In biology, selection is presumed to operate at the species, organism, as well as genome levels.

1.5. Other disciplines. Most other sciences seem not as permeated with optimality principles as economics, physics, chemistry, and biology. No doubt differences in the use of mathematical modeling explain much of this, although biology is not an especially mathematical discipline (for exceptions see Lotka 1956 or Barigozzi 1980). Whenever striving or competition is involved, the optimality perspective seems plausible. Nonetheless, sociology or anthropology generally do not claim (either quantitatively or qualitatively) that their societies, structure or processes are, in a general sense, optimal. Exceptions exist, however.

Wilson (1975), a sociobiologist, has used ergonomic theory to show that certain caste systems among social insects are optimal in their division of labor. In mature colonies, he argues, caste ratios approach an optimal mix in the sense of maximizing the rate of production of virgin queens and mates (given its size). Similarly, Becker (1976), an economist, developed formal models of marriage, discrimination, capital punishment, and so on, that rest on utility maximization. In psychology, signal detection theory, learning, operant conditioning, and choice theory have been formally expressed in optimization terms (Coombs et al. 1970). In sociology, in contrast, few formal optimization arguments are encountered, even though a flourishing subfield of mathematical sociology exists that is concerned with quantitative models of change and network structures (Coleman 1990; Sorenson 1978). Functional explanations in sociology and anthropology (Elster 1982) – such as social institutions serving the greater good – can also be construed as qualitative optimality arguments (e.g., Elster 1983 for a discussion of maximization in art).

Optimality arguments are also used in other physical sciences besides physics and chemistry. In geology, the paths of rivers have been shown to afford maximum throughput of water per time unit relative to the constraints of the terrain (Press & Siever 1974). In meteorology, weather systems are commonly found to equilibrate or dissipate in optimal fashion (Paltridge 1975). Last, cybernetics (Ashby 1956; Shannon & Weaver 1949; Wiener 1961) often models feedback and control systems in optimality terms. Furthermore, in the sciences of the artificial (Simon 1981), explicit attempts are made to design optimal systems. If the design problems are too complex for mathematical solution, analog models can sometimes be used to let nature solve them. For example, Kirkpatrick et al. (1983) used simulated physical annealing to optimize otherwise intractable combinatorial problems in computer design.

This section has established, at some length, the pervasive use of the optimality or at least extremal principles across a wide range of sciences. It is surprising, however, that most scientific texts fail to mention optimality in their indices. Whereas epistemological aspects of optimality have been extensively examined (Canfield 1966; Nagel 1953; Rosenberg 1985; Woodfield 1976), the concept is absent from the eight-volume index of the *Encyclopedia of Philosophy* (Edwards 1972), although it does contain a brief section on extremal principles. Optimality is also not mentioned, as a distinct philosophical entry, in *The New Palgrave: A Dictionary of Economics* (Eatwell et al. 1988). This paper seeks to redress this imbalance by asking scientists to clarify to what extent they deem optimality (in their own discipline) to be in the eye of the beholder as opposed to part of nature.

2. The concept of optimality

The use of optimality arguments in science involves a mathematical as well as an empirical component. The mathematical component is essentially a test of internal validity: Is the proposed solution indeed optimal (either locally or globally) relative to the criterion function and permissible domain? Although such questions can be complex, especially regarding the existence, uniqueness, and identifiability of a solution, the formal meaning of the optimality concept is well-defined. The real difficulty concerns the model's external validity: Does it correctly describe the phenomenon of interest? Three factors will be discussed as potentially troublesome concerning the empirical component: (1) Can anything be modeled as being optimal (given sufficient degrees of freedom)? (2) How comprehensive can our perceptions of nature be, given that they are species-specific constructions of reality? (3) Are nature's optimality principles, as uncovered across fields, mutually compatible?

Concerning the first point, the optimality concept can be easily trivialized as follows. Take any empirical law $y = f(x)$; express it as the first-order condition $y - f(x) = 0$; find an integrand $F = F(x,y)$ and boundary conditions such that $dF/dx = (\delta F/\delta y)(dy/dx) + \delta F/\delta x = y - f(x)$. Of course, the maximand F must be plausible, and so must the boundary conditions. The issue is by no means trivial (see Bordley 1983), however, especially in a field such as economics, where the permissible degrees of freedom concerning (1) the objective function, (2) the decision variables, and (3) the constraints are less well specified (e.g., unobservable budget constraints, transaction costs, information sets, cost of thinking, etc.). In the hands of a capable mathematical economist, a disturbingly large number of behaviors can be rationalized as being optimal, attesting to the dangerous and seductive flexibility of this heuristic. In Schoemaker (1982), such ex post facto use of optimality was referred to as *postdictive*, in contrast to predictive or positivistic models, which are falsifiable at least in principle.[4]

Similarly, in population ecology, foraging or mating behavior can easily be modeled as being optimal, which in turn has prompted various essays (Dupre 1987; Kingsland 1985; Maynard Smith 1984; Oster & Wilson 1978; Rachlin 1985) as to who is optimizing: the scientist or nature? Gould and Lewontin (1979), for example, particularly criticized panselectionists for their quick-footedness in proposing new selectionist explanations once old ones were discredited. Here is a sample of their criticism:

> If one adaptive argument fails, try another. Zig-zag commissures of clams and brachiopods, once widely regarded as devices for strengthening the shell, become sieves for restricting particles above a given size. . . . A suite of external structures (horns, antlers, tusks) once viewed as weapons against predators, become symbols of intraspecific competition among males. . . . The eskimo face, once depicted as 'cold engineered' . . . , becomes an adaptation to generate and withstand large masticatory forces. . . . We do not attack these newer interpretations; they may all be right. We do wonder, though, whether the failure of one adaptive explanation should always simply inspire a search for another of the same general form, rather than a consideration of alternatives to the proposition that each part is 'for' some specific purpose [i.e., the product of natural selection]. (Gould & Lewontin 1979)[5]

A second reason for being suspicious of grand optimality principles, such as Hamilton's least effort principle, Darwin's survival of the fittest,[6] or Smith's invisible hand (as a way of maximizing social welfare or at least allocative efficiency)[7] is that they were posited by creatures who themselves are part of the world they seek to describe (Whitehead 1920). Through our sense-awareness we presumably obtain just one of several possible representations of outside reality. In the case of vision, our normal range is limited from three to about eight thousand angstroms. And even within this narrow window we actively attend to less than 2% of the visual field. In addition, the eye is hardly an objective camera with cables, lenses, and screens (Hubel & Wiesel 1979), but part of a highly specialized information processing system that acts on prior categorization and expectations (Chomsky 1980; Hess 1973).

Our sense of there being just one reality (the one we all perceive) presumably stems from a uniformity within our species as to our mental primitives and pattern recognition.[8] Nonetheless, even our own limited window on the world is not always coherent. Optimal illusions remind us of the approximate nature of our perceptions. The notion that we can objectively perceive the surrounding world has proved untenable, especially in quantum mechanics (d'Espagnat 1979). In addition, nature has hardly proved to be commonsensical: Current conceptions of space (Callahan 1976) and time (Layzer 1975) are beyond most intelligent lay people and more than stretch our imagination. As Haldane (1927) noted, "The universe is not only queerer than we suppose, but queerer than we can suppose." Although researchers' bounded rationality (Simon 1957) is a caveat for all scientific theories, it especially applies to those claiming to have uncovered Nature's deepest or grandest principles.

The third concern is that optimality principles may be postulated which collectively do not add up to a coherent whole. Is the principle of least action, for instance, compatible with that of maximum chaos? (Most physicists would say yes, but not all; see Prigogine & Stengers 1984.) Can a natural selection principle, which operates reactively and with lags, ever lead to optimal adaptation in a changing world? Is evolution and the complexification of dissipative structures (Prigogine & Stengers 1984) compatible with the principles of entropy and chaos? Is *constrained* optimization (e.g., maximizing economic utility subject to a budget constraint) a contradiction in terms if the constraints can be relaxed (at a nonzero shadow price)?

What guarantees can we have that the infinite regression problem inherent in relaxing or (tightening) constraints at various metalevels will converge toward a stable solution (see Mongin & Walliser 1987)? How could we have argued that economic man was optimal 10 or 20 years ago, when by today's standards these past optimality models are simplistic and incomplete? Especially in the economics of information, the early optimization models made such strong assumptions as fixed search rules (Stigler 1961), whereas later models introduced "more optimal" variable rules (Rothchild 1973; 1974). Future generations will presumably consider our current optimality models unduly constrained and simplistic (Bounds 1987).[9]

Although the social sciences may suffer more from the fact that optimality is a moving target, conscious striving may be a condition favorable to optimality arguments. (Note, however, that consciousness and choice also open the door for suboptimal decisions.) [See also Libet: "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action" *BBS* 8(4)1985; and Searle: "Consciousness, Explanatory Inversion, and Cognitive Science" *BBS* 13(4)1990.] Animals are less deliberatively purposive (McFarland 1977; although see Griffin 1981), whereas plants and lower organisms are mostly passive participants in the process of natural selection. Nonetheless, they do undergo selection (by the environment). In contrast, physics, the most sophisticated of the optimality sciences, has the fewest a priori arguments in its favor. The laws of nature do not seem to have been selected for, nor do they appear mutable. Only appeals to a grand designer (God) or viewing optimality as a heuristic can justify the prominence of optimality theories in physics. To assess the extent to which physics' use of optimality involves teleology or metaphysics, Fermat's principle of least time is examined next.

3. The principle of least time

When we place a stick in the water, it seems that the angle above water is different from that below. This refraction phenomenon was extensively examined by the Greeks (especially Claudius Ptolemy) who constructed various tables of ingoing (or incident) and outgoing (or refracted) angles between air and water. However, it was not until the seventeenth century that an algebraic law was discovered (by the Dutch scientist Willebrord Snell) linking incident (θ_1) and outgoing (θ_2) angles. Snell's well-known law is $\sin \theta_1 = n \sin \theta_2$, in which n is a constant specific to the media involved. Although the task of science may seem completed when such laws are discovered, the French scientist Pierre Fermat (1601–1665) took it one step further (as is common in science; see Nagle 1961).

In going from A to B, he argued, light does not necessarily travel the path of minimum distance but rather that of shortest time (see Figure 1a). Suppose A is a point on the beach, and B a point in the ocean. What will be the quickest way to rescue someone drowning in the water at B when starting from point A? Ideally, one should angle the point of entry into the ocean (point X in Figure 1a) so that more time is spent in the faster medium (the beach), relative to the straight line route, and less in the slower medium (water). The optimal angle directly depends on the relative velocities in the two media, and corresponds to n in Snell's law.

Snell's law can be derived directly from the geometric puzzle discussed above. As light travels from point A to B in Figure 1a, its refraction at the border between air and water will be such that $\sin \theta_1 = (v_1/v_2) \sin \theta_2$. This, as proved in the Appendix, guarantees the shortest path in time. Fermat's ingenious principle generated a host of new hypotheses in optics, many of which proved correct. For instance, the least time principle predicts symmetry when reversing direction of propagation. It also predicts that light travels faster in air (v_1) than water (v_2), and that $n = v_1/v_2$. Thus, Fermat took Snell's law far beyond the original phenomenon of interest, deducing the shape needed for perfectly converging lenses as well as how light behaves with multiple lenses. No doubt Fermat's principle was very productive. But does it mean that nature optimizes?

The eminent physicist Richard Feynman expressed well the concern many feel when confronted with such principles as Fermat's. In his words:

> The following is another difficulty with the principle of least time, and one which people who do not like this kind of a theory could never stomach. With Snell's theory we can "understand" light. Light goes along, it sees a surface, it bends because it does something at the surface. The idea of causality, that it goes from one point to another, and another, and so on, is easy to understand. But the principle of least time is a completely different philosophical principle about the way nature works. Instead of saying it is a causal thing, that when we do one thing, something else happens, and so on, it says this: we set up the situation, and *light* decides which is the shortest time, or the extreme one, and chooses the path. But *what* does it do, *how* does it find out? Does it *smell* the nearby paths, and check them against each other? The answer is, yes, it does, in a way. (Feynman et al. 1964, Chapter 26, p. 7)

As Feynman explains, there is a quantum-mechanical view of Snell's law that gives considerable justification to Fermat's principle. Consider shining a flashlight into a rectangular water basin at an angle so that an underwater image appears on the back wall of the basin (as in Figure 1b). Why will the image be lower than expected from a straight line viewpoint (i.e., below point C)?

If light is viewed as photons (i.e., particles), the brightest image on the back wall occurs where the most photons strike. The probability of a photon striking is directly proportional to the number of pathways from point A (the origin) to some point B (on the back wall). Each pathway has a complex vector associated with it, whose angle is proportional to the travel time of that path. The overall probability of striking is obtained by adding all these complex vectors for a given point B and taking the squared length of the sum. The brightest point on the back wall will be that for which the various pathways are aligned in terms of vector angles as opposed to cancelling.

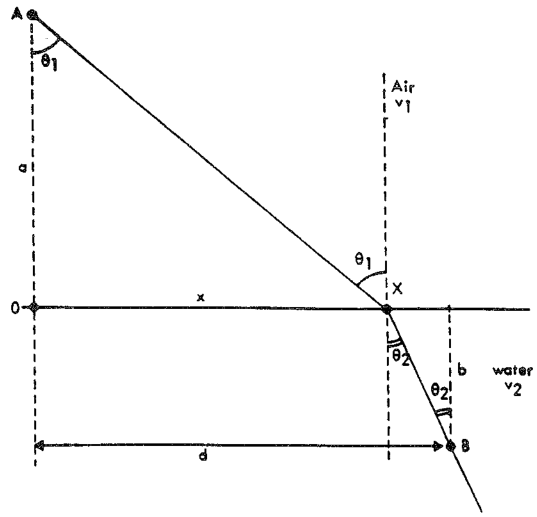

Figure 1a. Geometric illustration of Fermat's principle

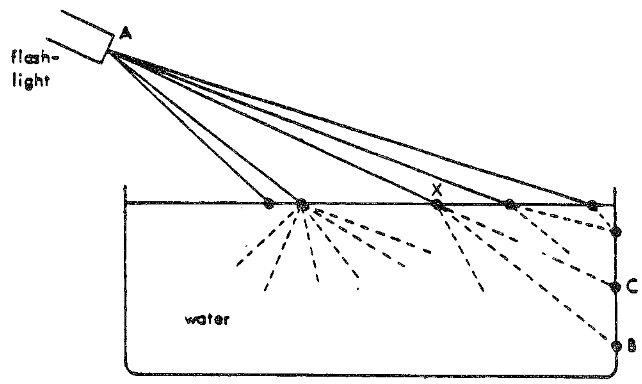

Figure 1b. Photon view of light refraction

The path of least time has the mathematical property that nearby paths will differ little from it in travel time, thus reinforcing the total vector length. In the language of the calculus of variations, the least time path is one for which first-order deviations from the path have only second-order effects on the time function. Only at a minimum or maximum do small deviations make no difference as a first-order approximation (e.g., in a Taylor series expansion).

Fermat was ingenious to have formulated a principle that accords so well with a deeper reality that was unrecognized at the time. His principle injected science with a metaphysical perspective, however. Whereas Fermat may have been inspired by the prevailing view that God designed a perfect universe, modern use of optimality stems more from a pragmatic (i.e., it works) than a religious belief. Nonetheless, the teleological nature of optimality arguments, especially in physics, sets them apart from causal theories. The next section contrasts teleological with causal explanations, and argues that both rest on psychological primitives that are only partly understood.

4. Scientific explanations

Of the many objects, organisms, behaviors, and systems around us, two classes might be distinguished depending on how we talk about them. If we attribute no inner purposes, needs, or intentions to them, they are nonteleological (e.g., a stone, fire, air, or stars). The other class, in contrast, is characterized by the presumption of intention, goal directedness, or plasticity (i.e., the ability to reach a goal from multiple directions or initial conditions). Humans, animals, and many lower organisms fall into this second, teleological class. The distinction is useful as it influences the types of scientific explanations deemed appropriate for each class. Causal explanations are typically expected for the nonteleological class, whereas intentional and causal ones are appropriate for the teleological class. Let us examine each type.

When asked, "Why did *x* occur," one acceptable answer is to identify a causal pathway from a set of plausible initial conditions to *x* (see Bromberger 1970). Many pathways (or causes) may exist, however. When a house burns down is it because of (1) a match being dropped, (2) the flammability of the carpet, (3) the failure of the sprinkler system, (4) the slowness of the fire department, or still other causes? Mackie (1980) defined an acceptable cause as being "an insufficient but nonredundant part of an unnecessary but sufficient condition." This view implies that causal explanations are usually not unique and depend on the expectational field of the observer. Objects and events in the world exhibit correlation, temporal antecedence, and contiguity (in space and time), but *not* logical necessity. The latter, as Hume (1888) emphasized, is a construction of the mind based on subjective conceptions and expectations. Indeed, Russell (1959) felt strongly that causality should be dropped entirely from our scientific dictionary. Nonetheless, nonteleological reasoning about scientific phenomena often rests on a primitive "causation" that is rather metaphysical and remains ill-understood in both psychology (Einhorn & Hogarth 1986) and philosophy (Bunge 1979; Davidson 1986; Kim 1981).

An alternative mode of scientific explanation is to ascribe purposes and intentions. In a sense, teleological explanations are the opposite of causal ones in that they reverse the temporal order. A proper cause, it might be argued, precedes its effect in time. When we say that birds build nests to lay eggs, however, a future (desired) state is alleged to govern the action. As such, the distinguishing feature of teleological explanations is not so much temporal order as the emphasis on goals and purposes (Wright 1976). Although purpose and intentionality may suggest notions of intelligence, consciousness, or even free will, the coexistence of physical determinism and purposiveness has become less problematic since Hobbes (1909) and Hume (1888) (see also Dennett 1978 or Schlick 1939).[10] Thus, teleological explanations (such as water seeking its lowest level) need not be metaphysical (see Woodfield 1976). [See also Dennett: "Intentional Systems in Cognitive Ethology: The 'Panglossian Paradigm' Defended" *BBS* 6(3)1983; and Schull "Are Species Intelligent?" *BBS* 13(1)1990.]

Teleological explanations seem especially appropriate when dealing with

1. cybernetic systems involving feedback (Rosenblueth et al. 1943),
2. homeostatic systems that try to maintain an equilibrium (Nagel 1953), or
3. ill-understood systems. [See Toates: "Homeostasis and Drinking" *BBS* 2(1)1979.]

Examples of each type are (1) a target-seeking missile, (2) a thermostat, and (3) animals, humans, or nature. In the latter case, the teleological approach often entails anthropomorphism. When a child is thrown over by a gust of wind, it will probably say the wind tried to hurt it (Piaget & Inhelder 1975). To children and adults, random or unexplained occurrences often assume personal meaning. Similarly, scientists may explain puzzling phenomena by imputing intentions. When asked why a balloon is round, we might say it chooses this form over others to equalize interior pressure. Such an explanation makes balloons lawful and predictable, while suggesting new avenues for exploration (e.g., what happens when you squeeze it or partly submerge it in water).[11]

Because of its metaphoric nature, the teleological approach can stimulate the mind to explore new avenues more effectively than the causal approach. As with causality, however, it rests on a primitive, namely, the concept of purpose or intention, which remains philosophically problematic (Canfield 1966; Chisholm 1956; Wimsatt 1972) and, according to some, unnecessary in our scientific lexicon (e.g., Skinner 1969). [See *BBS* special issue on the work of Skinner *BBS* 7(4)1984.] Psychologically, this primitive involves notions of forethought (i.e., a future orientation), assent (i.e., wish or desire), potency (i.e., the ability to affect the world), and consciousness of self (Warren 1916). Both causation and purpose appear to be psychological primitives that we naturally use to make sense of our surrounding world. It remains unclear which of these two crucial concepts is epistemologically more objective or valid (Braithwait 1966).

5. The rationality heuristic

Saying that people are rational, narrowly defined, implies that they use reason (*ratio,* Latin) to select means that maximize a well-defined end.[12] This assumption seems especially plausible in humans, who consciously strive to better their situation and must often do so in a competitive environment. This section examines both the pitfalls and promises of the optimality heuristic when applied to market behavior via the rational-economic man assumption. Economists generally assume that when the stakes are high enough people will act rationally for markets to reach equilibrium and be efficient. The argument rests on arbitrage opportunities being competed away by economic agents who possess (1) well-defined preferences, (2) comprehensive perceptions of the available options, (3) rational expectations concerning consequences (admittedly a thin theory of rationality; see Elster 1983), and (4) the ability to calculate which option has the highest subjective worth (e.g., expected utility). These premises are usually presumed to be "as if" assumptions and built into economic models to predict (rather than explain) real-world behavior. Several concerns exist, however, about the rationality heuristic (which is the primary optimality principle in social science).

First, rationality (in the sense of optimal behavior) is only operationally defined in limited circumstances. In the case of certainty or risk (i.e., known probabilities), near consensus exists among experts on what is meant by rational behavior. Under certainty, it means maximizing a utility function reflecting well-behaved preferences (i.e., those that are transitive, connected, etc.) subject to given constraints. Under objective risk, it is usually defined by economists as obeying the Von Neumann-Morgenstern (1947) axioms and maximizing expected utility (Keeney & Raiffa 1976). When outcomes are uncertain or ambiguous however (i.e., no agreed-upon probabilities exist), expert consensus breaks down and a variety of choice procedures are encountered, ranging from pessimism or regret minimization to *subjective* expected utility (see De Finetti 1975; March 1978; Milnor 1954; Schoemaker 1984; Shepard 1964).

In decision making under conflict, a similar absence of consensus is found (Luce & Raiffa 1957). Game theory offers numerous solutions, but only for the simplest of cases is a strong prescription found (such as the maximin principle, i.e., the notion that in certain zero-sum games one is best off pursuing a strategy offering the highest [max] return under the worst [min] possible countermove). In addition, group decision theory is plagued with a variety of impossibility theorems, making it difficult in most cases to identify the rational group decision (Arrow 1951; Satterthwaite 1975). Finally, intertemporal choice theory rests on a less solid axiomatic foundation than one would like. Discounted utility models, for instance, require Koopman's (1972) stationarity axiom, meaning that preferences are invariant under deletion or addition of time periods involving the same consumption levels. This axiom in turn forbids wealth effects and intertemporal synergies (see also Loewenstein 1987). Thus, although rationality may be well defined in the abstract, its operationality (in the sense of decisiveness) is often limited in scope.

A second concern is that for those instances where rationality is clearly defined, a growing body of laboratory and field studies suggests that people do *not* act according to the axiom (Einhorn & Hogarth 1981; Elster 1979; Hershey et al. 1982; Kahneman & Tversky 1979; Schoemaker 1982; Tversky & Kahneman 1981). Three defenses are usually invoked by economists concerning this counterevidence. One is that positive models need not be realistic in their underlying assumptions, as long as their predictions are accurate (Friedman 1953). Second, laboratory studies fail to provide the incentives and learning opportunities encountered in the real world. Third, in markets only a subset of the economic agents need to be rational (i.e., those trading at the margin or arbitraging) for the theory to work. It is usually not specified how many agents are enough, however, nor how quickly convergence to equilibrium should occur. Although the individual evidence against the rationality assumption is formidable, it remains an open question whether under aggregation (in real-world markets), the suboptimalities cancel, diminish or magnify (see Hogarth & Reder 1986). This question is especially crucial for such fields as policy or corporate strategy, where many of the interesting issues vanish when hyperrationality is assumed (see Schoemaker 1990).

A third concern regarding the rationality hypothesis is that it may be nonfalsifiable. The inconclusiveness of rationality principles in such crucial domains as conflict and group decisions, along with our limited understanding of the links among individual, firm, market, and macroeconomic behavior, permits a wide range of phenomena. In addition, such unobservables as transaction costs or the psychological cost of thinking introduce ill-specified frictional forces that allow considerable departures from the ideal. For example, Simon's (1957) theory of bounded rationality, according to which people satisfice rather than optimize, could be viewed as optimal once taking into account search costs, limited knowledge, and information processing limits. Of course, if all behavior can be argued to be optimal, the concept loses empirical content and falsifiability.

Nonetheless, the rationality heuristic has flourished in economics. In microeconomics it gave rise to equilibrium theories and theorems about comparative statics; in the macro realm it spawned rational expectations theory. Moreover, the rational economic approach has expanded beyond its traditional domain into animal economics (Battalio et al. 1981 [see also Rachlin et al.: "Maximization Theory in Behavioral Psychology" *BBS* 4(3)1981]; Mazur 1981), crime, marriage, and fertility (Becker 1976), public policy (Buchanan & Tullock 1962; Downs 1957), law (Posner 1973), conflict and war (Schelling 1960; Boulding 1962), and theories of organization form (Barney & Ouchi 1986). This wide scope speaks to the great power of the rationality heuristic. In the absence of detailed knowledge of the situation (e.g., about a firm or a marriage), ordinal predictions can be made (e.g., about direction of change) when varying certain parameters. This generality, however, is also the Achilles heel of economics; the precision and specificity of the hypotheses lessen as the theory's reach is extended.

6. The optimality heuristic and its biases

Given the wide use of the optimality heuristic, let us try to characterize its general nature and provide a brief summary of its potential biases. The following eight features characterize the optimality heuristic in general terms, with Fermat's example shown in parentheses.

1. *Posing a why question (or explanandum):* An unexpected or intriguing state of nature needs to be explained (e.g., Why is light refracted?).
2. *Bounding the domain of inquiry:* What are the problem's boundary conditions? What are the variables and what are their permissible ranges (e.g., light travels in straight lines; all refraction angles between 0 degrees and 180 degrees are *a priori* permitted)?
3. *Selection of salient features:* What aspect of the phenomenon can be anthropomorphized or examined via other metaphors (e.g., viewing light as traveling)?
4. *Teleological description of the system:* The phenomenon is modeled as seeking a desired end state, subject to certain constraints (e.g., light wishing to travel in the least amount of time).
5. *Search for the optimal solution:* Mathematical tools

are used to solve the optimization problem defined in step 4 (e.g., the time function is differentiated with respect to the angle of entry).

6. *Empirical comparisons:* The optimum solution obtained in step 5 is compared with the observed state in step 1 (e.g., Is the first-order condition that $\sin \theta_1 = (v_1/v_2) \sin \theta_2$ observed empirically?).

7. *Further refinement of the model:* If the predicted solution does not accord well with reality, the constraints or objective function might be adjusted to improve the model (e.g., under a very flat angle light may not refract but simply reflect).

8. *Generation of new hypotheses:* Does the teleological principle imply new predictions that can be tested empirically (e.g., What shape would be required for a perfectly converging lens?)?

The value of any particular optimality model depends in large measure on its plausibility (features 3–4) and the new insights it generates (features 7–8). If aspect 7 is highly contorted or arbitrary, for example, when the empirical phenomenon is force-fit into an optimality mold, the approach loses appeal and value. Whenever the underlying teleological principle is plausible and general (as with the law of least action), the associated metaphoric reasoning can be a powerful shortcut to new insights and solutions (Oppenheimer 1956). As is the nature of all heuristics, however, the optimality approach may be prone to various inferential biases (Kahneman et al. 1982). Not all of these are unique to optimality models, but each could seriously undermine their value, given the highly flexible nature of this heuristic.

6.1. Attribution bias. When scientific data fit some optimality model (ex post facto), it does not necessarily follow that nature or the agent therefore optimizes. To a large extent, optimality is in the eye of the beholder. It is the observer who is optimizing rather than nature (see Kitcher 1985). Nonetheless, in such fields as economics or ecology, agents' behavior is often deemed optimal whenever it can be accounted for by some optimality model. This, in my view, is a systematic and serious attribution error. As Ernest Mach (1883; see also Bradley 1971), Heisenberg (1955), and many others have emphasized, reality is the nexus between our mind and a presumed outside reality. The latter can hardly be understood independent of the observer (although see Schroedinger 1954; 1967).

6.2. Confirmation bias. Another inherent danger of the optimality heuristic is that its proponents may search more vigorously for confirming than disconfirming evidence. Falsificationism (Lakatos 1970; Popper 1968) would encourage such questions as, "How might we prove that light (or people) are suboptimal?" to appreciate better the limits of Fermat's (or economic) theory. In the case of light, we might ask how it would behave if it could travel in curved rather than straight lines, or if it minimized energy expended rather than travel time.

The confirmation bias may also slip in when reviewing the historical track record of the optimality heuristic. We remember well the names of Pierre Fermat, Charles Darwin, or Adam Smith, who were highly successful champions of the optimality principle. We hardly recall those who searched in vain for optimality or other aesthetically pleasing principles. One can only wonder what additional insights Einstein's formidable mind might have generated had he not stalked for more than 30 years the elusive unified theory of the (then) four basic forces of nature. Thus, the optimality heuristic may appear more successful in hindsight than a complete historical accounting would indicate.

6.3. Excessive rationalization. A third and related bias is that the optimality heuristic can result in tortuous rationalization. If one's prior belief in the value of the optimality heuristic is upwardly biased by the attribution and confirmation biases, it may seem productive to pursue it relentlessly. The attendant danger, however, is that the heuristic degenerates into the kind of thinking parodied in Voltaire's (1759) *Candide.* Reacting against the prevailing view that God had created a perfect world, Voltaire describes Candide's life as one disaster and mishap after another. Yet, amidst all the war, rape, famine, and suffering, each chapter reiterates that "this is the best of all possible worlds." By taking the argument to its extreme, Voltaire highlights the danger of Leibnitz's axiomatic commitment to a worldview steeped in optimality.[13] Economics, ecology, and sociobiology are some of the disciplines that have been criticized precisely because of their remarkable propensity to rationalize away anomalies.[14]

6.4. Illusion of understanding. A final important bias of the optimality heuristic is that it may create an illusion of understanding by describing rather than explaining. In Molière's play *Le Malade Imaginaire,* a doctor is asked to explain the tranquilizing effect of a drug. He tries to do so by attributing it to its "dormative faculty." Especially when phrased in Latin, such relabeling may instill an illusion of understanding (see Bateson 1979, p. 98). It fails as an explanation, however (as well as description), because it predicts nothing new and offers no further insight. Saying that light is refracted because it optimizes something is certainly more than relabeling (as new predictions were generated); however, it does fail to offer a process or causal account.

The positivist view that only prediction matters is fundamentally unsatisfying, and optimality principles consequently suffer from being too paramorphic. Does Fermat's principle really *explain* why and how light refracts? Do economic models predicting consumer reactions to price changes or equilibrium behaviors really explain how people behave? When ecologists argue that animals engage in optimal foraging, leaving one patch of land for another when the benefit/cost ratio gets too low, do they really explain how animals search for food? [See also Fantino & Abarca: "Choice, Optimal Foraging, and The Delay-reduction Hypothesis" *BBS* 8(2)1985; Houston & McNamara: "A Framework for the Functional Analysis of Behaviour" *BBS* 11(1)1988; Clark: "Modeling Behavioral Adaptation" *BBS* 14(1) 1991.] Each optimality principle, it seems, begs for an associated process explanation that describes causally, within the constraints of an organism or system, how it operates.

7. Conclusions

The optimality heuristic appears to be a very powerful principle of scientific explanation and inquiry. It is encountered in almost all sciences that are mathematical, and even in those that are not. Survival of the fittest, which is perhaps the grandest of all optimality principles, was formulated as a qualitative, conceptual cornerstone in Darwin's (1859) theory of evolution. Entropy and least action principles are other broad optimality laws, applicable to systems that do not overtly strive or compete (e.g., nonliving systems). Equilibrium notions and homeostatic behavior can also be interpreted as general optimality principles, covering wide domains of application. The issue arises, however, whether or not such optimality is solely in the eye of the beholder.

It was argued that for fields such as physics or chemistry the optimality principle is metaphysical unless viewed as a heuristic. As illustrated with Fermat's principle of least time, the refractive behavior of light can be viewed teleologically (i.e., light *choosing* the quickest route) or causally (i.e., photons following discrete paths that interfere or magnify). Although the causal view may seem more scientific, both modes of explanation rest on psychological primitives (i.e., the notions of causation and purpose) that remain ill-understood. Indeed, the teleological approach, owing to its metaphoric nature, can offer parsimonious summaries of physical laws while often suggesting new hypotheses. Especially in the physical sciences, the optimality approach (as a teleological principle) seems to have worked well (starting with Newton's laws of motion).

In the life sciences, the optimality heuristic has similarly been quite powerful. Formal applications are found in population biology, ecology, and medicine (e.g., mathematical models of the heart or knee). Because of natural selection, the justification for optimality principles seems stronger in the life than physical sciences. Nonetheless, adaptive systems can at best be optimal relative to a past condition (because of lags) and it will be hard to assess for optimality without explicit knowledge of the range of genetic variation. Furthermore, if random (i.e., nonadaptive) mutations are introduced, the range of potentiality becomes even harder to assess, and evolution may be neutral (Kimura 1979; 1983). Thus, the reactive nature of natural selection, the gradient climb toward local optima, and the unknown forward potential of mutations would appear to limit the appeal of optimality as an evolutionary principle. As emphasized by Jantsch and Waddington (1976), evolving systems are usually (1) imperfect, (2) in disequilibrium, and (3) unpredictable.

In human systems, however, a forward looking dimension is encountered. Humans are presumed to deliberate their future by learning from the past. Hence here the case would seem strongest a priori for the use of optimality arguments. Yet, with the exception of economics, the social sciences have hardly embraced the optimality heuristic (although see Zipf 1949). Not all of this can be attributed to the lack of mathematical modeling. In part it reflects a reluctance (possibly based on introspective evidence) to ascribe too much optimality to ourselves. As the anthropologist Eric Smith (1987, p. 205) claimed, "The bias in most of the social sciences [is] against reducing social institutions and processes to the action of self-interested individuals." Our inner complexity, as well as that of social aggregates, defies characterization in optimality terms without considerable simplification. Economics appears to be willing to pay the price of simplification; most psychologists prefer to render more detailed process descriptions (see Newell & Simon 1972; Sayre 1986) with fewer grand, unifying principles (such as utility maximization). Both approaches have their merit and reflect more than differences in mathematical sophistication.

Overall, the optimality heuristic has proved to be a powerful heuristic of science. Ironically, it is used most systematically and successfully in the physical sciences where its case is weakest a priori, and least in the social sciences (with the exception of economics), where its case is strongest prima facie (because of the conscious striving of people and the presence of competition and selection). The heuristic is powerful because it can offer an efficient summary of a system's behavior (Teller 1980) as well as suggesting new hypotheses. Its limitations, however, are that it can be too flexible, which may in turn lead to attribution errors, confirming rationalizations, and the confusion of prediction with explanation. Since its plasticity seems to be higher in the social than the physical sciences (owing to agents' consciousness, presumed free will and numerous unobservables), it is perhaps not surprising that optimality arguments are less eagerly embraced in the social realm.

The overall appeal of optimality arguments rests in part on our desire for simplicity and elegance in scientific theories. This was forcefully expressed by physicist Leon Lederman when he received the 1988 Nobel Memorial Prize for his work on subatomic particles. "My goal is to someday put [the basic laws of nature] all on a T-shirt . . ." he said. "We physicists believe that when we write this T-shirt equation it will have an incredible symmetry. We'll say: 'God, why didn't we see that in the beginning? It's so beautiful, I can't even bear to look at it'" (*Chicago Tribune* 1988). Symmetry, simplicity, and elegance appear to rank with optimality as among the most important driving forces in scientific inquiry (Chandrasekhar 1987). Our commitment to them, however, seems as much metaphysical as it does scientific, and is therefore in need of continual scrutiny.

In closing, I should emphasize that this paper is by no means an exhaustive treatment of the optimality heuristic. Given the broad range of issues involved, cutting across the physical, biological, and the social sciences, as well as philosophy, I humbly acknowledge my relative ignorance. Consequently, the ideas presented here should be viewed as personal reflections, aimed at inviting criticism and improvement. I leave it to each commentator and reader to decide what optimality really is: (1) an organizing principle of nature, (2) a set of philosophically unrelated techniques of science, (3) a normative principle for individual rationality and social organization, or (4) a metaphysical way of looking at the world. If the latter, we should strive to understand better when and how this root metaphor (Pepper 1942) enhances rather than obstructs scientific inquiry. As to my personal view, I consider the extremity principles encountered across sciences to represent both a common mathematical tool kit and a deeper assumption about nature's economy and elegance. My overall concern is that optimality principles reflect heuristic metastrategies for scientific inquiry that are prone to significant, often unrecognized, biases. The biases enumerated may be stronger in the social and biological sciences than the physical sciences because of the former's greater complexity and larger degrees of freedom.

8. Appendix

Derivation of Snell's law of refraction. Assume we start at point A and wish to reach point B as soon as possible with a travel speed of v_1 above the horizontal line in Figure 1a and a velocity of v_2 below this horizontal line. Our decision variable is x, the point on the horizontal border where we change relative velocities. To solve this problem, we must first define the travel time $T(x)$ as a function of x, and then differentiate with respect to x to find an extremum.

$$T(x) = \frac{AX}{v_1} + \frac{XB}{v_2}, \text{ where } AX = \sqrt{a^2 + x^2}$$

$$\text{and } BX = \sqrt{(d - x)^2 + b^2}$$

$$\text{or } T(x) = \frac{\sqrt{a^2 + x^2}}{v_1} + \frac{\sqrt{(d - x)^2 + b^2}}{v_2}$$

$$\text{Thus, } \frac{dT(x)}{dx} = \frac{2x}{2v_1\sqrt{a^2 + x^2}} + \frac{2(d - x)(-1)}{2v_2\sqrt{(d - x)^2 + b^2}} = 0$$

$$\text{or } \frac{x}{\sqrt{a^2 + x^2}}v_2 = \frac{d - x}{\sqrt{(d - x)^2 + b^2}}v_1$$

or $\sin \theta_1 = n \sin \theta_2$ with $n = v_1/v_2$.

ACKNOWLEDGMENT

Thanks to George Constantinides, the late Hillel Einhorn, Jon Elster, Eugene Fama, Victor Goldberg, Robin Hogarth, Paul Kleindorfer, David Lindley, George Loewenstein, Laurentius Marais, Stanley Martens, Merton Miller, Phillipe Mongin, Alex Orden, Alexander Rinnooy-Kan, J. Edward Russo, George Stigler, and William Wimsatt for helpful comments on earlier drafts.

NOTES

1. Scientists' high regard for theory and explanation (as opposed to classification or mere description) is clearly reflected in Luis Alvarez's remark about paleontologists: "They are not really very good scientists; they are really more like stamp collectors." Description, in my sense, occurs at the level of observation (e.g., Snell's refraction law in optics). Explanation involves reductionism or some appeal to other, possibly nonobservable, constructs (as in Fermat's principle of least time). Prediction concerns the ability to generalize beyond the original domain of inquiry in empirically correct ways. As Friedman (1953) argues, prediction need not involve description nor explanation. I sympathize, however, with those who dislike black box positivism (e.g., Samuelson 1963). Also, the earlier distinction between description and explanation is rejected by many (e.g., Bridgman 1948; Mach 1883/1942); however, see also Kuhn (1962) and Quine (1969).

2. More specifically, the currency in most optimal foraging models is the expected net rate of energy expended in foraging (Smith 1987). More complex currencies might be multidimensional (reflecting key nutrients), and include risk measures such as variance.

3. Immanuel Kant wrote: "In the natural constitution of an organized being, that is, one suitably adapted to life, we assume as an axiom that no organ will be found for any purpose which is not the fittest and best adapted to that purpose" (Kitcher 1987, p. 78).

4. Note that Boland (1981) in general refutes the falsifiability of utility maximization by resorting to Popper's (1968) view that "all and some" statements are not refutable. For a counterargument, see Mongin (1986).

5. Adapted from Gould & Lewontin (1979, p. 586) as quoted in Beatty (1987, p. 53) who critically reviews Mayr's (1983) counterargument that the strategy of trying another hypothesis when an initial one fails is common to all branches of science. The key issue, however, concerns the level at which the reformulation occurs and how many falsifications are needed before the underlying theory or paradigm is questioned. See also Kitcher (1985).

6. This phrase was actually Herbert Spencer's but captures well the spirit of Darwin's remarkable and sweeping theory.

7. The presumed optimality of the free market system has evolved much beyond Adam Smith's (1776/1976) insight that voluntary exchange is mutually advantageous and ultimately utility-maximizing for the parties involved. Coase (1960) emphasized that, under free exchange, property rights and resources eventually end up in the hands of those for whom they are most valuable (i.e., optimal allocation and usage) *independent* of initial distributions or endowments. This insight is in turn exerting an influence on the design and function of our institutions and our legal system (Posner 1977). For more detail see Hirshleifer (1987, Chapter 9).

8. Although sense organs differ markedly across species – with humans experiencing the world mostly through vision, bats or dolphins relying much more on hearing, dogs or rats using smell, certain fish and eels sensing electric fields, and spiders or scorpions registering primarily vibrations – Shepard (1987) has argued that all share the same three-dimensional Euclidean world in terms of internal representation. He hypothesizes that "our deepest wisdom about the world has long ago been built into our perceptual systems at a level that is relatively automatic and unconscious. If this is so, we may all be smarter than we 'think' – that is, smarter than our more recently acquired processes of articulate thought" (Shepard 1987, p. 267). This view further vitiates the notion that we could consciously comprehend nature's optimality, if any.

9. Many economists would argue that economic man is indeed optimal, but that our attempts to model economic man are not, and thus undergo continual improvements (analogous to physicists' improved models of the atom). The paradox here, for economics, is that these bright scientists acknowledge their own suboptimality in the fiercely competitive game of model building while attributing perfect rationality to their lesser brethren in the game of economic life.

10. Lenoir (1982) offers a fascinating insight into teleological versus mechanical views among nineteenth century German biologists when Darwin's ideas gained currency. Although the Darwinian view won out, biological language is still replete with teleological terms, from selfish genes to survival instinct.

11. When plunged into water, the balloon will change in size and shape, but not instantly. As physicist David Lindley noted: "Air inside the balloon will bounce around and oscillate until the new stable shape is assumed. The significance of this is that the bouncing around represents a physical and entirely causal way by which the air in the balloon literally tries out other configurations before reaching the one dictated by least energy" (personal communication, Nov. 9, 1988).

12. A broader definition would include the rationality or moral worth of the ends as well, and would perhaps reduce the emphasis on the use of explicit reasoning (see Elster 1989).

13. I am referring here to Leibnitz's principle of perfection, according to which "God selects that universe for which the amount of perfection is a maximum" (Rescher 1979, p. 26).

Leibnitz viewed this principle as a moral rather than a logical necessity, stemming from the choice of the best.

14. Recent candidates in financial economics include attempts to rationalize dividend policies of firms (Miller 1986), and to explain away high volatility of stock prices (Kleidon 1986; Shiller 1981) including the stock market crash of October 1987 (Roll 1989; Malkiel 1989). In the area of preference theory, interesting rationalizations can be found regarding utility theory anomalies (Cohen 1981; Loomes & Sugden 1982; Machina 1982; Schoemaker 1982), and the so-called preference reversal phenomenon (Grether & Plott 1979; Slovic & Lichtenstein 1983). Of course, whether these are "true" rationalizations or standard scientific defenses is largely a matter of opinion.

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Optimality and human memory

John R. Anderson
Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA 15213-3890
Electronic mail: *anderson@psy.cmu.edu*

A preoccupation with optimality explanations can undoubtedly lead a scientist to miss crucial insights into a particular class of phenomena. This is true of any single-minded approach toward theorizing, however. It is equally true that neglect of optimality considerations can blind us to certain insights. As Schoemaker notes, optimality analyses have not been much developed in psychology and I (Anderson 1990a) have argued that this has been at considerable loss of insight. Nowhere is this truer than in our understanding of human memory. I would like to use the domain of memory to look at the issues that Schoemaker raises in a different light.

On a first impression, commonsense basis, human memory might seem a poor candidate for optimality explanations. We are forever complaining about how bad our memories are, failing to provide us with the facts we need. Modern information-processing theories of memory have taken this first impression to heart and produced theories of memory mechanisms without concern for their optimality characteristics. The work in this field illustrates that the falsification problems Schoemaker raises are not unique to optimality theories. For example, the question of whether there is a separate short-term memory has been addressed for decades without any adequate resolution, and it is now regarded by many as nondecidable. The field also shows that there can be progress despite such problems. We certainly know a great deal more about the phenomena surrounding short-term memory or working memory (Baddeley 1986).

Just as Schoemaker is ill at ease with optimality theories that do not explain how the optimization is achieved, I have long been ill at ease with the lack of concern about why the memory mechanisms we have proposed behave the way they do. Marr (1982) complained about a similar exclusive concern with mechanism in theories of perception. Marr's discussion of the computational theory of perception applies equally well to memory: "An algorithm is likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism in which it is embodied" (p. 28). What is the problem that human memory is trying to solve? Memory is basically trying to perform an information-retrieval task in which it must retrieve from a data base of millions of experiences those memories relevant to the situation at hand. This is essentially the same task that modern computer information-retrieval systems try to perform (Salton & McGill 1983). Compared to the performance of these computer systems, human memory is far less nonoptimal. Thus, we see that first impressions about the plausibility of an optimization explanation can be deceiving.

To explore the possibility of an optimality explanation of memory, we needed a theory of the information-processing demands placed on memory by the environment. Adapting theories developed for library borrowing (Burrell & Cane 1982) and file access (Stritter 1977), we (Anderson & Milson 1989) showed that many memory phenomena could be seen as optimal responses to the statistical structure of retrieval requests from the environment. More specifically, we showed that human memory displays the fastest retrieval latencies and highest probability of recall for the information that is statistically most likely to be needed.

This research was based on information-retrieval demands placed on nonhuman systems, however. More recently, Lael Schooler and I set out to study carefully the actual information processing demands placed on humans. We looked at a number of such computerized sources of input to humans as topics in the *New York Times*, electronic mail messages, and words spoken to young children. This is not the place to describe our results in detail, but in every case we found that memory functions mirrored perfectly the statistical properties in the environment. Thus, we are finding, contrary to all expectations, that human memory seems exquisitely tuned to the statistics of information presentation in the environment. Whether this will ultimately be viewed as a case of optimization or it will be given some other explanation, it illustrates the potential for optimality considerations to lead to novel insight. This would never have been known unless we chose to penetrate beyond the apparent nonoptimality of human memory.

The domain of human memory also casts light on the question of how such optimization can be achieved. The actual proof of the optimality of memory requires relatively sophisticated Bayesian mathematical analysis. There has been no claim that humans do this, however. In Anderson (1990a) I show that these computations can be performed by neural activation mechanisms that are exceedingly simple by today's standards for neural computation (McClelland et al. 1986; Rumelhart et al. 1986). Thus, we see that one can get dumb mechanisms to behave "as if" they are optimal.

Finally, memory phenomena illuminate the question Schoemaker raises about the lag of the system: All the domains we looked at are definitely modern in their character (*New York Times* stories, electronic mail messages, modern parents' speech to their children). Now if our memories evolved to be optimal they presumably did so in the presence of rather different input in the past, yet we find memory optimized to the statistics of the present. We also see that they would behave optimally if they were libraries or file systems. The obvious conclusion is that some universals of information retrieval do not change much with time or material. It is just such universals that we would expect our memories to internalize.

Perhaps human memory is the perfect contrast to economics in terms of evaluating the quest for optimality. One can argue that in economics there is such a fixation on optimality analyses that crucial insights have been missed. In the case of human memory, however, optimality considerations have been ignored at the cost of insight. There is no general answer to the question of whether the quest for optimality is a positive or negative heuristic for science. Sometimes it works and sometimes it does not, and we have to find where and how it works by engaging in the empirical inquiry that is science. [See also Anderson: "Is Human Cognition Adaptive?" *BBS* 14(3) (forthcoming).]

Optimality as an evaluative standard in the study of decision-making

Jonathan Baron
Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104-6196
Electronic mail: *baron@cattell.psych.upenn.edu*

I want to try to extend Schoemaker's excellent article by considering the use of normative models as evaluative standards (rather than as descriptive models) in the study of decision-making and related activities.

As Schoemaker points out, many departures from normative models, such as expected-utility theory and Bayesian probability theory have been discovered in recent years (Baron 1988). For example, Kahneman and Tversky (1979; 1984) showed the following kind of "framing effect," in which the option chosen depended on irrelevant features of the way a choice is described to subjects: When subjects are told that 600 people are expected to die in an epidemic, they prefer saving 200 to a $\frac{1}{3}$ chance of saving 600 (and a $\frac{2}{3}$ chance of saving none), but other subjects prefer a "$\frac{1}{3}$ chance that nobody will die (and a $\frac{2}{3}$ chance that 600 will die)" to the certain death of 400. Subjects presented with both descriptions would easily recognize that a $\frac{1}{3}$ chance of saving 600 is the *same* as a $\frac{1}{3}$ chance that nobody will die, just as saving 200 is the same as the death of 400. Apparently, subjects compare outcomes to a reference point (none saved or none dead) that is easily influenced by the form of presentation.

The logic of such a research program might be described as follows. We begin with a model of optimal decision-making. We look for systematic, consequential errors defined as departures from the optimal model. Then we ask whether people can somehow be induced to improve their performance, to become more optimal (e.g., Baron & Brown, in press; Larrick et al., in press). If the model of optimal decision making is correct, then we can infer that we are doing good through such efforts (*ceteris paribus*).

This logic depends on the correctness of the optimal, normative model. We cannot rely here on behavior to establish correctness, for it is behavior that we are evaluating. We might try to rely on intuitions that we have on reflection, either before or after these intuitions are systematized into a coherent philosophical theory of rationality. A danger in this approach, however, is that whatever is causing the errors in behavior will also be reflected in our intuitions (as argued by Spranca et al., in press), even when they are systematized. For example, many theorists feel compelled to include a term for ambiguity in their normative model (Ellsberg 1961), despite strong arguments against such a move (Frisch & Baron 1988).

I have argued (Baron 1988) that normative models are best thought of as derived from an analysis of the general goals already inherent in the situations at issue. In decision-making, our goal is to maximize the extent to which our goals are achieved, that is, to maximize utility. From this criterion, expected-utility theory can be derived. For example, one form of the important *independence axiom* is that our choice should not depend on any outcome that will be constant across our options in a given uncertain future state. If option *A* leads to outcome *X* in state *S1* (e.g., the coin is heads), option *B* leads to *Y* in *S1*, and both options lead to Z in *S2* (tails), then our choice should depend on a comparison of *X* and *Y*. *Z* does not matter. This principle follows from the fact that the achievement of goals is not affected by events that do not occur. If *S1* occurs, then the comparison of *X* and *Y* is what matters. This comparison cannot be affected by the nature of Z, because *S2* did not occur. If *S2* occurs, our goals are achieved equally by either option, so the nature of Z is again irrelevant. (We assume that *X* and *Y* are not affected by the *knowledge* of the nature of Z. That is, *X* and *Y* are full descriptions of all relevant aspects of the outcome, including psychological aspects.)

To the extent that this kind of argument holds, we can conclude that such violations of the normative model as those discovered by Kahneman and Tversky lead to suboptimal achievement of goals. At issue then is whether there is some *method* of decision-making that would consistently lead to improvements in the choice of options, judged by the standard of the normative model.

Note that this use of normative models is distinct from a weaker kind of evaluative criterion, namely, *consistency* between decisions and rules of decision-making that the decision maker consciously accepts. By this criterion, a person who does not accept expected-utility theory (or the axioms that imply it) cannot be judged irrational for failing to conform to it. Such a criterion of consistency is not a true use of optimality. Inconsistency alone does not imply any failure to achieve one's goals, for the principle that one adopts may be nonoptimal even when one's action is optimal. It is not clear that helping people become consistent will do them any good (as pointed out to me by Deborah Frisch).

A second kind of consistency condition is illustrated by the epidemic problem. An inconsistency is found here between different decisions, not between a decision and a principle. Such inconsistency does imply that at least one decision is not optimal, assuming that different decisions cannot both be optimal. But inconsistency alone is a weak heuristic for discovering the nature of the irrationality or for finding ways to correct it. Fortunately, the epidemic problem can also be understood as a specific kind of departure from expected-utility theory.

In summary, I have pointed to a concept of optimality that does not suffer from the criticisms that Schoemaker correctly makes of other optimality concepts. The idea of a normative model as an evaluative standard can, in principle, help us achieve our goals.

Optimality as a mathematical rhetoric for zeroes

Fred L. Bookstein
Center for Human Growth, University of Michigan, Ann Arbor, MI 48109
Electronic mail: *fred_l._bookstein@um.cc.umich.edu*

Schoemaker says that the optimality heuristic can "offer an efficient summary of a system's behavior as well as suggest new hypotheses" (sect. 7, para. 5). I disagree: Optimality principles are not efficient summaries but at best tautologies. The proper semantics of "optimality" is none of the four choices the target article offers us in its last paragraph, but instead: a mathematical/rhetorical style for reporting zeroes of derivatives generated by observation or by artifact.

As the target article explains in section 1, the path a mechanical system takes is that about which the variation of the action vanishes. This offers "excellent numerical results for otherwise intractable problems," Feynman et al. (1964) noted. But the article never returns to the actual (not the meta-) physics of this example. Lanczos (1970, p. 77) explains that analytical mechanics rests on one single postulate: "The virtual work of the forces of reaction is always zero for any virtual displacement which is in harmony with the given kinematic constraints." That is, mechanical systems, whether static or dynamic, are always in virtual equilibrium – derivatives are zero – with respect to the forces restricting them. The vanishing of the variation of action (Hamilton's principle) is a logical consequence of this exact constraint (*op. cit.*, p. 92). Optimality principles like least action are not in fact used to explain anything. They are only mathematical reformulations of the postulate. Although they allow the solution of many textbook problems, they do not augment our understanding of nature.

Schoemaker hints at this same critique at the end of section 6, noting that any optimality principle begs for an associated process explanation. Only in physics is this logical hiatus closed. The path taken by a ray of light in the world is no more an "optimization" than the nearly elliptical orbits of planets around the sun. In holding constant the function of position and velocity appropriate for motion driven by an inverse-square central force, ellipses "optimize" nothing; they merely encode the law of gravity in a more directly observable form. Contrary to Schoemaker's assertion (sect. 7, para. 2), what drives mechanics and light is geometrization (Lanczos 1970, p. 139), not optimization. The situation in physics would be identical with "trivialization," the overinterpretation of which is properly criticized (for economics) in the second paragraph of section 2 – except that, in physics, Hamilton's principle holds exactly, so that problems get solved.

In physics, then, optimality is fundamentally a mathematical strategy, a reference to an integral having variation zero in place of a quantity at equilibrium. Elsewhere it should be characterized not as a scientific principle but as a rhetorical style. Outside physics, quantities can be claimed to be optimal only when their derivatives, with respect to some interesting quantity (such as time), are zero. An "optimal principle," although an impressive form for such reports, is neither explanation nor summary of such data; it is only a restatement in different notation. Any explanation or prediction associated with these observations requires some mechanism to account for the invariance. There are many different modes of this accounting – none, in my view, very promising, mainly because the rhetorical alternative does not lead to any useful suggestions about measurement.

The power of Hamilton's principle in physics derives not only from the exact law that it embodies by integrals but also from the ability of the physical scientist to measure many different aspects of the same mechanical system to comparable precision. After the mechanicist has expressed the identical Lagrangian function, the identical "action," in a new coordinate system, he is often capable of actually measuring the values of those generalized coordinates in independent ways, as the mass of Mars is measured by the periods of its satellites, for instance. Using different instruments, one can measure position, time, velocity, and acceleration with astonishing, indeed "unreasonable," consistency and precision (Wigner 1967).

Outside physics and chemistry, this direct measurement of integrals with precision equivalent to that of integrands appears flatly impossible. In sect. 2, para. 2, Schoemaker claims that the maximand $F(x,y)$ that integrates an exact law $y = f(x)$ must be "plausible." He ignores a far more crucial requirement: F must be explicitly measurable independently of f, like energy vis-à-vis motion or bending of light around the sun vis-à-vis gravitation. Otherwise the integration to optimality is mere mathematical rhetoric without any empirical consequences. But this "otherwise" includes all the biological, psychological, and social sciences.

Perhaps the commonest source of inappropriately "optimal" quantities in science is an observer's preference for reporting the values of quantities that do not change. A psychometrician, for instance, prefers to study IQ as an "underlying factor" that can be postulated not to vary by age: an invariant over a universe of virtual changes in test composition. A tautology from factor analysis transforms the purported invariance (stability against substitution of test items) into an optimality (of explained covariance). In this way, entirely by artifact, IQ becomes a construct that optimally "explains" performance on its subtests. A similar adulteration of science by optimization is the invocation of parameters of linear regressions – optimizing (least-squares) fits to observed data – as if they meant something other than the so-called normal equations, the equations of decomposition by paths (Wright 1954) that underlie the usual formulas. An average, for instance, is least-squares because the sum of deviations about it is zero – because the data balance there. Unless we can explicitly measure that sum of squares, the relevant property of the average is not this optimal integral but the zeroing. Nor is the magnitude of a correlation evidence of "optimality" of anything: Sums of squares are not independently measureable (except in physics).

Schoemaker asks his commentators to clarify "to what extent they deem optimality . . . to be in the eye of the beholder versus part of nature." I would counter that optimality is in the mouth of the beholder, not the eye. When optimal principles help us solve numerical problems in the physical sciences, it is because they are mathematically equivalent to the exact constraints that in fact apply throughout the universe. In all other sciences, the failure of optimality principles to lead to knowledge derives from the absence of exact constraints and observable zeroes of derivatives: The optimality heuristic does not lead to new knowledge because it does not lead to new measurements. From neuropsychology through economics (see McClosky 1985) and beyond, there seem to be no invariants available on which the mathematical rhetoric of optimality can usefully work. Outside physics, whenever scientists properly attend to the quantities that they claim to be measuring, and to the evidence that such quantities are invariable, the "explanations" embodied in the language of optimality will be found irrelevant to the understanding of empirical patterns. I am thus much more pessimistic than the author about the potential of this "heuristic" for advancing reliable knowledge.

The quest for plausibility: A negative heuristic for science?

R. W. Byrne
Scottish Primate Research Group, Department of Psychology, University of St. Andrews, St. Andrews, Fife KY16 9JU Scotland
Electronic mail: *pss10@st-andrews.ac.uk*

Schoemaker's argument that optimality principles cannot be any more than heuristics boils down to the demonstration that these principles have worked well in sciences where their justification is implausible, yet they have been found less useful just in those fields where they should – he believes – apply most. This appeal to plausibility is explicit: "The value of any particular optimality model depends in a large measure on its plausibility," and the notion of plausibility used is whether the entities the science deals with might be imagined to *want* an optimal solution. Thus, in the physical sciences, which "lack the conscious striving and forethought characteristic of humans," Schoemaker deems it a priori unlikely that optimality should have much mileage, so he is surprised how useful principles like Fermat's have been. Whereas the social sciences "have conscious striving as a condition favourable to optimality arguments," even in biology, the plausibility of optimality is seen as plausibility of purposes: "Animals are less deliberately purposive, whereas plants and lower organisms are mostly passive participants in the process of natural selection."

Plausibility may seem a plausible heuristic, but consider the statements:

(1) The balloon wants to equalize its interior pressure.
(2) The plant wants to maximize its rate of gene survival.
(3) The baboon wants to manipulate the other to support it.
(4) The baby wants its mother's comfort.
(5) The businessman wants to get rich.
(6) The lecturer wants the student to understand.

No scientist has problems with statements (1) and (2), which are obviously teleological shorthand for more laborious non-teleological explanations (though this is culture-specific: Some people find "My birth wants me to be happy" an entirely literal description of fact).

Most ethologists and psychologists consider it quite possible that baboons or babies *may* have purposes of the kinds implied in (3) and (4); but whether they *do* is hard to test, and quite irrelevant for understanding most of their behaviour – including whether an optimality principle such as natural selection has moulded their behaviour to what we observe. Thus they treat the statements as teleological shorthand for other types of explanation, except in those (albeit very interesting) cases where the mental processes of animal or child are specifically at issue. The problem of whether the lags inherent in natural selection make it likely that a current form is "ideal" at the moment or still slowly evolving toward an ever-changing optimum, is a quite unrelated issue, clearly easier to resolve for a relatively static genus like *Nautilus* than for *Homo*.

It is certainly plausible that businessmen might want riches, but this has little bearing on whether their behaviour obeys an optimality principle, let alone whether their behaviour is "a good thing" for anyone else, or the economy. A teacher is perhaps more likely to close on an optimum style more quickly, the firmer the intention is to communicate – but other factors may have greater importance than mere wanting. Indeed, it is far more likely that a plant will optimize, since the very flexibility and adaptibility of human behaviour makes nonoptimal acts slower to be weeded out. Schoemaker is aware of this, and to some extent his dilemma is a straw man. But if some people in the social sciences are *really* getting in a muddle between the literal fact of wanting an ideal and the teleological shorthand of optimality principles, then perhaps it would be helpful to try a temporary ban on *all* literal meanings of mental state terms, in favour of treating them all as if in quotes, even in cases where conscious striving towards an end is plausible or provable. This, I suppose, is what radical behaviourism tried to do for psychology – except that behaviourists forgot to mention the "temporarily" qualifier!

Criteria for optimality

Michel Cabanac

Department of Physiology, Laval University, Quebec G1K 7P4, Canada
Electronic mail: *cabanac@lavalvm1.bitnet*

The word optimal is derived from the Latin optimum, the best. The word is therefore comparative and implies a reference. All the problem of choosing between "optimality, a principle of nature" versus "optimality is all in your mind" lies in the frame of reference. One may consider, as pointed out by Schoemaker, that all of science is a set of conventions, and hence when we discern a general principle of nature it is also an artificial construction of the human mind. What is at stake is no less than the nature of mind and of science. Yet one should not question again and again the reality of the outside world, and whether what we perceive is reality or illusion. The problem has been solved by Descartes. Let us take it for granted that we think within boundaries. Now, within these boundaries, the concept of optimality must obey the rules of science. We must have in our possession criteria of optimality. When we decide that a system maximizes a variable to a value that we declare optimal we must be able to discern a finality for the observed maximization.

Schoemaker's skepticism about optimality is welcome when optimality is used as a *petitio principii*, which is often the case in ethology. Yet there are cases where the circularity of the concept of optimality can be opened, where the word refers to an identifiable criterion. Biology and economics provide examples where optimal behavior can be measured against reproductive and physiological fitness or against financial efficacy. Without such a criterion the optimality of a system remains a mere description, as seems to be the case when it is used in physical sciences.

When male dungflies sit copulating with females long enough to maximize the number of eggs fertilized, and short enough to allow time to search for new females (Parker 1978), their behavior may be considered optimal from the point of view of population genetics. When subjects placed in a conflict of motive – fatigue versus cold discomfort – maximize sensory pleasure and thus both thermoregulate and avoid tachycardia (Cabanac & LeBlanc 1983), one can consider their behavioral choice as optimal from the point of view of physiology. When a broker chooses an investment with maximal return and minimal risk, this behavior can be considered optimal from the point of view of the client's income.

On the other hand, I fail to see in the maximization of entropy in a steady state or in a parabola a finality homologous to the survival and profitability found in the above examples from biology and economics. In this biologist's possibly blind eye, entropy, steady state, and parabola are nothing more than descriptions. I fail to see which nonsubjective criteria can be applied to the optimality examples selected by Schoemaker in the realms of physics and chemistry. In these sciences one can indeed define metaphysical criteria of optimality as seen from an eschatological point of view but it should be made clear at the same time that this is not science. Schoemaker suggests that there might be a general organizing principle of optimality in nature, but in accepting such a principle one also leaves the strict realm of science.

Some optimality principles in evolution

James F. Crow

Genetics Department, University of Wisconsin, Madison, WI 53706
Electronic mail: *wrengels@wiscmacc.bitnet*

Schoemaker's target article points out a number of ways in which optimality principles have been used in science. I would like to follow his lead and mention a few more, from evolutionary theory.

The most popular general quantitative statement in evolution is R. A. Fisher's (1930/1958, p. 37) fundamental theorem of natural selection. In his words, "the rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time." Fitness is the capacity to survive and reproduce. Fisher's preferred measure of fitness is the intrinsic rate of increase (which he calls, appropriately, the Malthusian parameter, *m*), the solution of the equation

$$\int_0^\infty e^{-mx} l_x b_x dx = 1$$

in which l_x and d_x are the probabilities of surviving to age x and of reproducing at that age. The genetic variance (now usually called the genic or additive genetic variance) is the variance of the least squares estimates of the genotypic fitnesses. Fisher likened his principle of increasing fitness to the increase of entropy in a physical system.

It is obvious that evolutionary progress depends on genetic variation; it is not so obvious that the relevant metric is the genic variance. The theorem's importance, or at least its fascination, is evidenced by the flood of discussion it has elicited. Fisher's elegant opacity leaves room for a seemingly endless string of interpretations of what he really meant. There are also extensions to make the theorem more exact and more comprehensive. It is too much, with all the friction and noise of biological systems (especially with Mendelism), to expect a formulation that is at once simple, exact, and complete. If treated as a good approximation, however, Fisher's theorem packs a great deal of evolutionary insight into a simple statement. For the rate of change of a trait correlated with fitness, such as a performance character in livestock, the genic variance is replaced by the genic covariance of the trait with fitness (for a recent article see Nagylaki 1989).

More relevant to Schoemaker's discussion is the theorem in integrated form: When selection moves a population from one state to another, the difference in the fitness (or correlated character) of the two states is equal to the genic variance (or covariance) summed over the time required to move from one state to the other.

But does the population follow an optimum path? The answer is yes. Svirezhev (1972) has provided such an interpretation by finding the appropriate Lagrangian. Since this work was published in Russian, it is not widely known in English-speaking countries. Consider a single locus in which the i-th allele has frequency p_i. Svirezhev integrated the quantity

$$\sum_i \frac{\dot{p}_i^2}{p_i} + \sum_i p_i(m_i - \bar{m})^2$$

over the path from t_1 when the population is in state 1 to t_2 when the population is in state 2. The superior dot indicates the time derivative, and m_i is the mean fitness of all genotypes containing the i-th allele. Svirezhev showed that the population follows a trajectory so that this quantity, when integrated along the path, is stationary. The first and second terms can be roughly compared to kinetic and potential energy.

What is important from the evolutionary standpoint is that, for a randomly mating population, this quantity is also Fisher's genic variance. The second term is obviously half the genic variance (the factor 2 comes from diploidy). For the first, note that

$$\dot{p}_i = p_i(m_i - \bar{m})$$

so the first term, too, is half the genic variance.

We arrive, with Svirezhev, at the statement that natural selection operates in such a way that the path followed as a population changes from one state to another is the one that minimizes the total genic variance over the path. Evolution gets the biggest fitness bang for the variance buck.

If this principle had been discovered in an earlier century it might have been regarded as evidence of an all-wise, optimizing Creator. Today it is placed on a lower pedestal, but it has (to me, at least) great appeal. Just as one can obtain Newton's laws from the principle of least action, one can find the equations of gene-frequency and fitness change from Svirezhev's principle.

Recent work in evolutionary theory has usually dealt with such more limited, but more concrete subjects as evolution of the sex ratio and stochastic theories of molecular evolution. Some examples are mentioned in Schoemaker's article. The "least variance" principle has not so far had the heuristic value that its counterparts in physics have had. Yet, its elegance has aesthetic appeal and it provides another, possibly insightful way of looking at evolution.

Natural selection doesn't have goals, but it's the reason organisms do

Martin Daly

Departments of Psychology and Biology, McMaster University, Hamilton, Ontario, Canada L8S 4K1

Electronic mail: *daly@mcmaster.bitnet*

It seems to me that Schoemaker misconceives the unifying principle of the life sciences when he equates natural selection with goal-directed striving. Darwin himself perpetrated such a vitalistic analogy when he made "nature" an agent that "selects," but Darwin understood full well that the selective process has no goals of its own. It is instead the reason entities that do have goals exist. Schoemaker never quite succeeds in articulating a fundamental distinction between living and lifeless systems: Purposive ("teleological") concepts are *properly* applied to organisms because they have goal-seeking processes instantiated in their structures as a result of the evolutionary process, whereas the application of such concepts to lifeless phenomena always entails an analogy to organisms. The consequences of biological phenomena constitute a legitimate and essential part of their explanation: What they achieve is in a concrete sense why they exist. The same cannot be said of lifeless phenomena. Contra Schoemaker, it was not thanks to Hobbes or Hume that "the coexistence of physical determination and purposiveness has become less problematic." It was thanks to Darwin.

Immediately after referring to optimality models of sex allocation, foraging, and mate choice, Schoemaker concludes, "In biology, selection is presumed to operate at the species, organism as well as genome levels." This is incorrect and betrays an incomprehension of the "levels of selection" arguments in biology. All the authors cited in this section reject species-level selection as a source of adaptation. Moreover, the "organism" and the "genome" represent the *same* "level of selection"; what is most commonly at issue is whether selection is more usefully construed as taking place between whole genomes (organisms) or between components thereof (Cosmides & Tooby 1981; Dawkins 1982).

Schoemaker proceeds to note that "sociology or anthropology generally do not claim (either quantitatively or qualitatively) that their societies, structure or processes are, in a general sense, optimal." But it is not that "functional" theories have rarely been advanced in these disciplines. Rather, sociological and anthropological functionalisms have generally failed miserably, for reasons transparent to a Darwinian: There is no rationale for expecting societies to function as if they had "purposes," because there is no process that could plausibly provide them with the sort of integrated "designs" and goal-directedness that selection gives to individual organisms. An individual organism manifests integrated purposiveness because the fitnesses of its constituent genes and coreplicons (Cosmides & Tooby 1981) strongly covary and have been selected to "cooperate." A society, by contrast, lacks integrated purposiveness because its constituent organisms routinely achieve fitness at one another's expense. Those who postulate adaptive functions of society are seeking functionality at an inappropriate level of the hierarchy of life (Williams 1966).

Schoemaker cites Gould & Lewontin's (1979) notorious critique of "adaptationism" with evident admiration, and demands "how many falsifications are needed before the underlying theory or paradigm is questioned" (Footnote 5). The answer is that the falsification of particular adaptationist hypotheses does not and should not call adaptationism into question. The failure of the hypothesis that the heart is an organ of cognition was no reason to abandon the search for its real function.

Adaptationism, the paradigm that views organisms as complex adaptive machines whose parts have adaptive functions subsidiary to the fitness-promoting function of the whole, is today about as basic to biology as the atomic theory is to chemistry. And about as controversial. Explicitly adaptationist approaches are ascendant in the sciences of ecology, ethology, and evolution because they have proven essential to discovery; if you doubt this claim, look at the journals. Gould & Lewontin's call for an alternative paradigm has failed to impress practicing biologists both because adaptationism is successful and well-founded, and because its critics have no alternative research program to offer. Each year sees the establishment of such new journals as *Functional Biology* and *Behavioral Ecology*. Sufficient research to fill a first issue of *Dialectical Biology* has yet to materialize.

As a "second reason for being suspicious of grand optimality principles," Schoemaker notes "that they were posited by creatures who themselves are part of the world they seek to describe" with species-specific sensory, perceptual, and cognitive constraints. Our limitations in these regards constrain all of science, and they constrain unscientific ways of knowing, too. Schoemaker offers no rationale for considering this ubiquitous problem to be especially a nuisance with respect to optimality arguments.

Organisms, scientists and optimality

Michael Davison
Psychology Department, University of Auckland, Auckland, New Zealand
Electronic mail: *mcdop@ccu1.aukuni.ac.nz*

Scientific explanations. Everything a human or animal or star or stone does is rational – it has a reason, even though this reason may be rooted in chaos. This is the working hypothesis of science. We may understand these reasons in some sense: We may be able to state something about the precursors of that behavior and be aware that those events often precede the event in question; we may know something of the way the system is constructed and that the event to be explained has a reasonably high frequency of occurrence under a variety of conditions; or we may be able to point to a likely future event that will probably follow the event to be explained. All these are explanations of some sort – respectively, historical, structural, and teleological ones. Each of these, to be able to function as an explanation, represents a high correlation between something and the event to be explained. A pigeon pecks at a grain because it has not eaten recently, because it is hungry, and because it wants to eat. Each of these may be a sufficient explanation for some purpose and for some audience.

I would like to look at these three sorts of explanations, and to judge them on their completeness for explaining behavior, and on their mutual independence. First, the teleological explanation.

If I explain an animal's behavior in such terms as "it wanted to eat," I seem to be making an indirect statement about the current state of the system: It is a hungry system. This is perhaps sufficient to tell me that it is likely to do *something* to eat, but not specifically what it will do on the way to eating. What specifically is done depends on the sort of system it is and on the current environment in which it finds itself. So the teleological explanation is, I suggest, a partial, indirect structural explanation of the current state of the organism. A teleological explanation can stimulate the mind (sect. 4), but it also has another property: that of satisfying the mind. Satisfaction through partial explanation is surely not what we need.

Teleological explanation has a further problem. Although it may embody a summary of some of the knowledge that we as scientists need to have about the current state of the system at the time the behavior is emitted, the explanation cannot easily be used to predict the future occurrence of the behavior because, when it is used, it is a post-hoc explanation – at least until an explanation of this type can be sustained at a very high level. Optimality explanations purport to have reached this level and so can be used in a predictive way.

As suggested above, teleological explanations are incomplete, and we also need to know something of the structure of the system – what system it is, and what its history is. One could argue that the historical data are unnecessary because a full optimality account could take a system from an arbitrary starting point and predict the trajectory it would take over a substantial time period while it approached some stable state in some predictable environment. Unless the system has a very long memory (for instance, being chaotically sensitive to initial conditions), historical data may become available within the terms of the explanation. The final argument might be that the *principle* of optimality is independent of the sort of system involved (light and water, birds and food) and of what is optimized (time traveling, rate of intake). So optimality, as a high-level teleological explanation, could be complete in principle, and also in application, given that we are clear about the structure of the system under investigation.

But is it right? The problem is the example given in the target article in section 3 – an asymptotically simple example, quite unlike any situation that must be accounted for in the behavioral sciences. The light ray must have "smelled" the water from a distance and been ready to change direction instantly. It can't be wrong, it can't be slow to react, the experimenter (or nature generally) can't surprise the ray, it can't forget, it can't learn. It is truly an optimal system. The explanation in terms of minimizing time can fit what happens perfectly, but it simply cannot be what happens. The problem is in the minimizing time *from A to B* – the ray *knows* where *B* is before it has left *A*. I will return to similar problems later.

As an explanation, the current state or structure of the system, if one could take a snapshot of it, could in principle do a good job of predicting at least the immediate likely future behavior of the system. Such an explanation is in principle complete, though in practice, because of technological limitations, it is likely to remain incomplete. The completeness in principle exists for one reason only – the current state of the system, through various memorial processes (which may in general extend back through many generations, but may also strongly represent recent occurrences) is an encapsulation of the system's history – in behavioral terms, the history of the organism and of the environment in which it lives and has lived. It is in the merging of these twin histories that explanation ultimately resides.

I suggest therefore that explanations in terms of the future and the present both ultimately reside in the past, and it is there we must look for explanations that are both in principle complete and technologically feasible (at least to a reasonable accuracy).

I have above approached the problems of explanation from the point of view of a behavioral scientist, which is quite different from the approach that might be taken by a classical physicist. The physical sciences have developed and blossomed by dealing effectively with very simple systems compared with those that are the subject matter of the behavioral sciences. It is, indeed, only since the acknowledgment of chaos that the physical sciences have begun to be confronted with system behaviors that are similar to the starting point and irreducible data of the behavioral sciences. Behavioral scientists have, I believe, a rather better appreciation of the promises and pitfalls of the varieties of explanation than do most physical scientists. Behavioral scientists appreciate the disastrous effects of teleological explanations of behavior provided by the media. It is all too difficult to convince a public fed such pap that there are real questions about the causes of behavior and of events generally.

Optimality in the behavioral sciences. Let me start by saying that, so far as I know, assumptions that animals or humans will behave so as to optimize some currency in their transactions with their environments have not been sustained with any generality. For example, in signal-detection theory, the classical assumption that animals and humans behave so as to maximize their payoff probabilities does not describe the subjects' behavior (e.g., Hume & Irwin 1974; McCarthy & Davison 1984). Rather, animals and humans behave nonoptimally. In the study of animals foraging in patches in which there may or may not be a prey item, animals generally stay longer than the time that would maximize their rates of prey intake. And in simple choice situations, animals change neither their rates of performance nor their distributions of choices between alternatives in a way that maximizes their rates of food intake (e.g., Davison & Kerr 1989; Vaughan & Miller 1984). If animals are rational, then either rationality does not imply optimization, or animals are not being rational about things that *psychologists* think they should be rational about. We can add the traditional epicycles: The subjects are optimizing something else; we need to include the costs as well as the payoffs of behavior; we need to include the costs and payoffs of leisure; we need to consider the type of economy in which the subjects are working, and so on. Yes, the optimality approach has stimulated research (sect. 4), but I think that the long-term value of such hypothetico-deductive cuts and thrusts is doubtful, and begins to resemble the celebrated and sterile Hull-Tolman controversy. Often those of us who are more interested in empirically based research are asked by journal editors and reviewers to take a stand for or against particular theories, because it seems that everyone is constrained to stand somewhere on this polygon. This rather strains the idea of "stimulating research."

Having taken this empirical stand against a blinkered optimization approach (and having been told by one of the developers of signal-detection theory that testing the theory was inappropriate because data were irrelevant to the theory itself), I must now stand back and ask, "What is an optimality approach?" – or, perhaps, "What is *not* an optimality approach?" The light ray described in the target article is an example of an omniscient, instantaneous optimality. Wherever the ray starts, it already "smells" all its future and "knows" all its subsequent *B* locations. It has no desire ever to arrive at *C*, and there is no way we can ever change its preordained path. It tries, always successfully, to minimize its lifespan. The sun to my left has already smelled my spectacle lenses, and its rays already know when I will turn my head. If this is optimality, then it happily contains within it, already, a description of my future behavior *and* my future environment. Either this is terribly important, the final unifying insight of all science, or it is silly.

In the behavioral sciences, a rankly optimal approach has a number of characteristics that can be seen as degradations of the light-ray case.

(1) some rationally derived function, simple or complex, is maximized or minimized;

(2) this maximization may be subject to various constraints;

(2a) constraints on the discrimination of the independent variable values, either singly or severally, either independently or interactively;

(2b) constraints on the output of the organism in the same ways;

(2c) constraints on the time and history of the organism's functioning – the window or time horizon over which the maximization will take place. This will determine, for instance, whether the organism can smell distant peaks or will "satisfice" on local hills.

(2d) the variance in the environment over time, in conjunction with (2c), will constrain molar optimality, and depending on the above constraints, including sensitivity to variance and to trends, will constrain molecular optimality. The light ray is affected by none of these constraints. It is these constraints that will allow an optimality theorist to satisfice on suboptimal experimental predictions, at least at an early stage in the development of such a theory.

Other theories are couched in less obvious optimality terms. For instance, according to melioration theory (Herrnstein & Vaughan 1980), an animal subject to constraints will reallocate time spent between available alternative choices so as to equalize the local rates of goods obtained from each alternative. This is an optimality theory because the subject minimizes the difference between local rates of goods. As another example, let us entertain a theory about animals foraging in patches that may contain a prey item. Maybe it gives up on a patch when the hyperbolic value of the delayed items that may remain falls below the hyperbolic value of traveling to a new patch (this theory does rather better, at least in some cases, than does maximizing overall prey-rate theories). Is this new theory an optimality theory? Again, yes, because the animal's behavior changes so as to equalize (or minimize differences between) two or more continuously calculated variables, subject to constraints.

By this line of reasoning, I come to the conclusion that all theories can be interpreted, in one way or another, as optimality theories. The equality sign (or any other relationship sign) brings with it a possible optimization interpretation. I suspect, therefore, that optimality approaches are absolutely fundamental to all quantitative science – both as a metaphysics in giving the illusion of understanding, and as a heuristic, providing some motivation and organization for enquiry. But optimality approaches are just one of many ways of providing a metaphysic and heuristic, and thus they are not unique. I guess the major problem for me is when optimality approaches are used normatively and as a kind of behavioral commandment, I feel this is the way I should behave to get the most out of life. If theories are used in such a way, then I should soon be able to buy an Optimality Calculator to which I can input my current choice alternatives and calculate, for example, my optimal tennis stroke (the example used by Herrnstein 1990), and maximize the probability of my winning a game. The fascinating question is whether or not I will *feel* happier about the outcome. I suspect not: Despite the objective evidence that my game had improved significantly, I think I would still harbor the belief that I could do better, probably by reverting to my previous style. Such questions should also be asked about the machines that determine, via optimality models, whether a nuclear counterstrike should be ordered, and how it should "best" proceed. The results may be optimal for the machines and the theories, but are they optimal for humans?

The goal of science is to predict the future output of systems. This was relatively easily attained in the classical physical science of idealized systems, but the science of chaos, be it physics or behavior, requires system and environmental history for explanation. Simple dissipative systems – e.g., a pendulum with friction – predictably move to simple minimized states, and can be described as optimizing. But the dynamical systems of animal behavior or chemical reactions can come to what is known as a *detailed balance* at many and varied points, and they are often subject to dependence on initial conditions. To search behavior for explanations armed only with optimality is not an optimum strategy for understanding.

Vaulting optimality

Peter Dayan[a] and Jon Oberlander[b]

[a]Centre for Cognitive Science, Department of Physics, and [b]Human Communication Research Centre, University of Edinburgh, Edinburgh EH8 9LW, Scotland

Electronic mail: *[a]dayan@cns.ed.ac.uk and [b]jon@cogsci.ed.ac.uk*

The target article provokes three comments and one more general criticism. First, the reason it is not surprising that "[the optimality heuristic] is used most systematically and successfully in the physical sciences . . . and least in the social sciences" is not its increased plasticity, as is claimed. Rather, it is the greater accuracy and wider applicability of mathematical models in the physical sciences. As Schoemaker amply points out, the equations governing many physical dynamical systems can be viewed through the smoky glass of optimising principles. But in the social sciences it is rather easy to provide myriad examples to refute a hypothesis with substantial quantitative rather than just qualitative import. Optimality hypotheses, dealing in quantities, are hard to come by because quantitative hypotheses in general are hard to come by.

Second, equilibria and optimality bear a more complex relationship to each other than is revealed by viewing the former as just "minimising a difference function defined on the actual and ideal states," as suggested in the context of chemical equilibria. For example, Maynard Smith's (1974) evolutionarily stable strategies (ESSs), which are indeed equilibria, might often be better viewed as suboptimal, in terms of different criteria.

A third point is that the issues the target article highlights are slightly obscured by the absence of a distinction between the messages that might be better directed at general explanation in science and those that are specific to optimality principles. For example, of the eight features of the optimality heuristic, only the fourth, "teleological description of the system," is really confined to optimality; all the others seem to be perfectly general. "Confirmation bias" is also not restricted to this heuristic.

The more general criticism can be seen clearly in a paradigmatic environment for the optimality heuristic – cognitive science. Humans are entities for which both teleological and causal explanations may genuinely invoke processes of optimisation, whereas with water or exchange rates, only the teleological explanations trade on optimisation. As seen in the target article, optimality assumptions are rife at the "higher" cognitive levels, for instance in postulates concerning rationality; but they have also been made about the "lower" subpersonal levels, for example in postulates concerning energy minimisation for constraint satisfaction. Unfortunately, only the significance of levels of explanation and description is ever alluded to. The claim is made, for instance, partly in the context of economic explanation, that "each optimality principle, it seems, begs for an associated process explanation that describes causally, within the constraints of an organism or system, how it operates." Surely this confuses the levels.

For concreteness (rather than correctness, Foster 1990), consider Marr's three levels (Marr 1982). At the computational level, the task a system performs is described and possibly justified on the grounds of appropriateness; at the algorithmic level, representations commensurate with the task and the algorithm by which it is carried out are defined; and at the implementational level, the precise physical realisation of the algorithm is described. As Fodor teaches (Fodor 1975), the fact that psychology has an independent existence at all is a function of the different modes of theoretical explanation at these different levels. The target article suggests an unhappiness with a computational-level optimality principle unless its algorithmic and/or implementational level are also evident. This is unlikely to be a fruitful methodological restriction.

In this context, questions about the use of optimality should be directed at the computational level. How felicitous is it to suppose that human cognition is optimising some measure? The quick-footedness Gould and Lewontin (1979) note is also evident in the discussion of planning under the assumptions of bounded rationality, as seen in a dispute between Dennett and Fodor (Dennett 1987; Fodor 1981). Fodor criticises Dennett for being too wedded to predicting others' actions on the basis of assumed rationality, meaning rationality in the narrow sense defined in the target article. Dennett responds that rationality is inevitably bound in terms of the time and space available for processing, and it is therefore appropriate to predict assuming these bounds. But this threatens to make the notion of optimisation too trivial to be of value.

In drawing our attention back to the uses and abuses of optimality, the target article raises a number of important hurdles that users of optimality principles should vault. Needless to say, cognitive science is "tripping" happily.

Optimality and constraint

David A. Helwega[a] and Herbert L. Roitblat[b]

Department of Psychology, University of Hawaii at Manoa, 2430 Campus Road, Honolulu, HI 96822

Electronic mail: *[a]davidh@uhccux.bitnet; [b]herbert@uhccux.bitnet*

Optimality theories have found widespread application in diverse fields such as physics, economics, and ethology. The value or validity of using an optimal decision theory to analyze animal behavior (including human behavior) periodically comes into question. Investigators have raised a number of criticisms opposing optimality as an ecological criterion. Among these are attacks on the rationality assumption of optimal decision theory and the capability of animals to perform the necessary computations for behaving rationally.

There seems to be some confusion about the nature of optimization. As Schoemaker notes, the concept of optimality does not refer to perfection in the sense of an achieved goal, it refers rather to preference for alternatives that maximize the relevant currency. In biological systems, inclusive fitness is the ultimate currency. Natural selection is perforce an optimizing function in that it selects from a set of alternatives those that are more successful at reproducing their genetic copies. Natural selection ranks genotypes in order of their fitness, and selects those that are most fit. An individual that failed to maximize its fitness would quickly (in evolutionary time) be replaced by an individual that was more successful at maximizing its fitness.

Although inclusive fitness provides a powerful argument for the use of optimality as an explanatory factor in animal behavior, specific analyses frequently have to use such a more proximal currency as energy or risk of predation. The use of a proximal currency is a simplification. It does not deny that the real currency is fitness, it merely acknowledges that this currency may be too complexly related to the specific behaviors under consideration to make useful predictions. This simplification is for the benefit of the investigators, rather than a strong assumption about the factors that control the animal's behavior. As scientists we can consider only a few variables at a time. Of these, our tendency is to select (at least initially) those that are most salient and that correspond to the largest differences in the phenomena we are investigating.

Another important confusion concerns the relation between the optimum as a goal and the process of optimization as a mechanism. For example, one argument is that "populations may spend more time tracking moving fitness optima (climbing adaptive peaks) than they do sitting at the summit optima" (Cody 1974, in Pierce & Ollason 1987, p. 113), a criticism that is repeated in the Conclusions of the target article. This criticism conflates the existence of optima in a landscape with the process of optimizing. Optimizing is the process of approaching a maximum (or minimum); it is not the maximum (or minimum) itself. A phenotype that does not perform as well as possible (e.g., by satisficing) will lose to a competitor that does. This is as true for an animal that is competing against other animals as for one competing against the environment (Maynard Smith 1978), an artificial dichotomy that obscures the fact that game theory is an optimal decision theory. The key point is that an optimizing animal is performing as well as it can, subject to constraints (Roitblat 1982). Organisms are subject to many constraints, including structural ones introduced during their evolutionary history (Gould & Lewontin 1979), limited capacity to obtain and process relevant information (Kamil 1978; Kamil & Roitblat 1985; Krebs et al. 1978), locomotor limitations, competing demands (Houston & McNamara 1987), and many others. An unconstrained (e.g., omniscient) animal would experience infinite fitness. Infinite fitness, however, is obviously unattainable. For example, an animal with infinite gonads might be able to produce an infinite number of offspring, but the animal's inability to move might subject it to infinite predation, thus negating the benefit of gonad size. The maximally fit animal will (*inter alia*) compromise between gonad size and mobility to maximize its reproductive success.

A major criticism of the use of optimal decision theory is that the models are too complex computationally to be used by any real biological organism. This criticism stems from two misconceptions. The first is that the complexity of deriving a model is confused with the complexity of the process that is being modeled. The second is that the mechanisms animals use to compute solutions are different from the computational mechanisms used by mathematicians. For example, Houston and McNamara (1988) provide a thorough review of the use of dynamic programming to induce the decisions an animal should make if it were acting in an optimal manner. The computational complexity of dynamic programming is known to rise exponentially as more variables and more states are added to a model. As a result, some readers of Houston and McNamara felt that the magnitude of the computations performed by the simplest of dynamic programs was not within reach of many animal species (cf. Clark 1990). This point is entirely irrelevant to use of optimal decisions by the animal, however. The complexity of dynamic programming is a difficulty that the investigator has in deriving a prediction of what the animal can be expected to do. Dynamic programming is an algorithm used by investigators; it is not, and cannot be, a model of how the animal performs that computation. For example, computing the trajectory between two moving objects can be a computationally intense problem involving the solution of multiple differential equations. When characterized in this way, such computations would seem to be beyond the capabilities of most animals, but many dogs are adept at catching flying frisbees, because they have a nervous system that automatically and implicitly computes such solutions. Dogs do not need to know differential calculus to catch frisbees and foraging animals do not need to know dynamic programming to select optimal decisions.

Recent developments in artificial neural networks (e.g., Wasserman 1989) suggest ways that animals could compute the solution to highly complex problems by using only very simple computational mechanisms. Artificial neural networks are massively parallel systems that contain complexly interconnected networks of very simple computational elements or conceptual neurons. [See Hanson & Burr: "What Connectionist Models Learn: Learning and Representation in Connectionist Networks" *BBS* 13(3)1990.] These elements receive inputs over weighted connections, and produce outputs as a function of the weighted sum of these inputs. Although each element in the network has very limited computational capacity, a three-layer network can compute any arbitrary function (Hecht-Nielsen 1987). Artificial neural networks have been derived that rapidly compute the solution to such highly complex constraint satisfaction problems as the travelling salesperson (TSP) problem (e.g., Kirkpatrick et al. 1983). We have designed a simple neural network that simulates an individual fledging decision, as described by Clark (1990, sect. 5) for murres. We generated a dataset that contained the variables body weight, day of season, and the decision (fledge or remain in nest) from the equations provided by Clark. Each case represented the fledging decision of one murre that managed to survive at least until the end of the simulated season. The data were presented to a backpropagation neural network that consisted of two input neurons (weight, day), a two-unit hidden layer, and one output neuron (fledging decision). The network successfully learned to compute a fledging decision based on these examples of previous successful fledging decisions in a way analogous to that by which evolution could select (over multiple generations) networks with the appropriate connections and weights. The point of this demonstration is that a simple network, consisting of very simple elements, and working with past successful examples, is sufficient to solve a problem previously modelled using a complex dynamic programming algorithm. Moreover, network modelling of this sort has the potential for widespread application rather than being constrained to a given species. The network may or may not be useful in deriving predictions for the animal's optimal behavior (we think such models can be useful), but it does show how such solutions could be computed. Computational complexity may thus provide a constraint on our ability to develop models of the animal's behavior, but it is not likely to be particularly relevant to the animal's ability to solve optimization problems.

We agree with Schoemaker's hesitation in accepting the evolutionary or ecological validity of the specific forms of some optimal decision models. We do not share his reluctance to accept optimization as a principle when dealing with animal behavior, however. Optimal decision models are hypotheses about the usefulness of certain variables in predicting animal behavior. Behavioral ecology investigations are not ordinarily directed at tests of the optimality assumption, per se; rather, they are directed at assessing the validity of the variables hypothesized to be the controlling factors. To recognize the importance of constraints on optimality is not to deny the usefulness of optimality theory. To argue for optimality is not synonymous with panselectionism or perfectionism. Animals do the best they can (i.e., optimize) with the information that they have and within the limits of their ecology and structure.

Types of optimality: Who is the steersman?

Michael E. Hyland

Department of Psychology, Polytechnic South West, Plymouth, Devon PL4 8AA, England

Electronic mail: *@prime-a.poly-south-west.ac.uk*

It is important to distinguish between two types of optimality, one that is *not* based on control theoretic principles and one that is. Consider, first, optimality that is not based on control theory. A lake has a relatively constant depth because the rate of flow out of the lake is a function of the depth of the lake and hence of the rate of flow of water into the lake. The lake does not "try" to have a constant depth; nor was it designed in that way. The constant depth is the consequence of the laws of nature, which are a primitive or set by God. Whether God sets these laws out of concern for lakes is uncertain. On the other hand, a thermostatically controlled room heating system keeps a room at a relatively constant temperature because the system is designed by people, not God, to detect deviations from a person-determined optimum and elicit behavior (i.e., heating) that minimizes those deviations.

Schoemaker (sect. 6, para. 3) writes, "To a large extent, optimality is in the eye of the beholder." In the case of optimality without control (for example, the assertion that the lake has an "optimum" depth), such optimality is entirely in the eye of the beholder. Many of the examples given by Schoemaker, for instance, natural selection, fall into this category. Organisms do not seek an optimal form – it is a law of nature that the less optimal are more likely to die. The law determines that which is optimal not vice versa. It *is* possible to suppose that evolution has a purpose, but if that assumption is made, then the purpose is that of God (e.g., de Chardin 1968). If one does not accept divine intervention, then such optimality is the consequence of humans recognizing constancy in outputs of lawlike systems.

On the other hand, human-made control systems (it is unclear what Schoemaker [sect. 4, para. 2] means by distinguishing cybernetic from homeostatic systems) have reference criteria (or set points or goals) that can be described as optimal states and that are the preferred states of people. This second type of optimality is not in the eye of the beholder, but in the eye of the person who constructs the system and consequently in the public domain. The set point of a thermostat is indicated on a dial, and the reference criteria of a robot are described in technical manuals.

Some researchers, myself included, believe that a person operates as a hierarchically organised control system (Carver & Scheier 1982; Hyland 1988; Powers 1978). If that is the case, then a person's reference criteria are optimal states, but they are optimal states that are part of the person rather than part of the outside observer. Where do people's reference criteria come from? They do not come from other people, as in the case of a robot. And few would argue that they come from God. It seems that people have the capacity to create their own reference criteria, or optimal states, and that these are theoretical rather than in the public domain. That is an important difference between the reference criteria of robots and those of people.

The scientist (who is an outside observer) needs to make inferences about people's reference criteria (which are theoretical constructs), and this inferential process may be wrong. Indeed, the rationality heuristic (Schoemaker, sect. 5) is based on the assumption that people have a type of reference criterion that they may *not* have or one from which discrepancies may not produce appropriate action. The rationality heuristic may be in the eye of the beholder for a quite different reason: because it is wrong (e.g., Evans 1989).

Here is a final point about the relationship between teleological and causal explanations. Some time ago, Lewin (1943) argued that purposive explanations are causal because behavior is caused by the organism having a representation of future events prior to the behavior. Behavior is not (and cannot be) caused by the future events themselves. Thus, if reasons are used in Lewin's sense, the explanation is achieved, not by providing a rational account of the behavior but by actors being supposed to have had (from their own perspectives) a rational account of that behavior. In terms of time perspective, control theory is similar to Lewin's purposive explanations: Purpose is explained in terms of representations of the future. The optimal state, goal, or reference criterion is a steersman (from which the word *cybernetics* derives) that guides action to where the steersman wants to go. The fundamental question raised by Schoemaker's quest for optimality is, "Who is the steersman – God, man, or machine?"

Natural science, social science and optimality

Oleg Larichev
Institute for Systems Studies, Prospect 60 let Octjabrya, 117312, Moscow, 9, USSR.

I find the target article to be good and interesting, but I do not agree with Schoemaker on some problems.

In analyzing the application of optimality principles in different research fields, Schoemaker creates a fuzzy boundary between two different classes of problems: (a) Problems for which it is possible to find reliable models (those that allow one to make reliable predictions and that generate repeatable results), and (b) problems for which the models depend on evident or hidden assumptions made by the researchers. This boundary assigns problems in natural sciences to the first class and divides into two parts the problems in economics.

Optimality principles in the natural sciences (and partly in economics) are very convenient techniques for describing the objects under study; their utility in this respect does not need further support. The success of optimality principles in explaining and predicting events speaks for itself.

Temporary attractiveness on the part of some techniques can be found very often in science; we have no universal techniques. Our knowledge of the environment is also limited; we can have relatively good models for some problems and relatively bad ones for others: For example, current models for global climatic changes are quite unreliable; in the future they will be more reliable.

The biases cited by Schoemaker (sect. 6) pertain only to economic problems. The example of Fermat's principle (sect. 3) merely demonstrates normal scientific development: Scientists usually make guesses or propose the hypotheses; eventually one finds a rigorous theoretical basis for confirming (or rejecting) these guesses.

Things are quite different in economics. For one class of problems there are the attempts to construct "objective models." For example, Wagner (1969) proposes "the objectivity principle." Two researchers working on the same problem must derive the same model. In such models human freedom of behavior is eliminated; people are subordinated to the logic of the objective model (e.g., Wagner 1969, the transportation problem). (By the way, optimality principles are dominant in operations research.) In economics, however, there is a wide class of problems in which people do have freedom of behavior, and many results from psychology and sociology demonstrate the limited nature of human rationality.

We now have two "rationality camps" (Jungerman 1983), with the followers of classical rationality theory and utility theory providing their respective supporting arguments (cf. Schoemaker, sect. 5) But lately there have been more and more confirmations of the limits on human rationality. For example, human behavior in real life problems of multiattribute classification (Larichev & Moshkovich 1988) cannot be explained by Multi Attribute Utility Theory (MAUT) or Subjective Expected Utility (SEU). We see instances of limited rationality not only in laboratory tasks performed by the students, but in decision makers' behavior.

Yet there are significant differences among the ways that limited rationality demonstrates itself: The performance of experts and decision makers rarely gives rise to contradictions (as that of students does); experts usually simplify complex problems in dramatic ways.

There are two reasons why the rationality heuristic has flourished in economics. First, the means of testing it in human behavior are often connected with averaging over many decisions, the majority of which are quite simple (one alternative is dominant or quasidominant). The second reason is mentioned by Schoemaker (sect. 5): It is always possible, for example, to find explanations of human behavior in terms of SEU theory. But predictions of human behavior with the SEU model are very unreliable and even subjective.

It is now time to touch on the very base of utility theory, which provides a foundation for a theory of rational behavior. Many of its underlying assumptions can now be called unjustified.

First let us consider measurement. Von Neumann and Morgenstern (1953) assumed that an analogy exists between measurement in the natural sciences and one in economics. In their words:

> As to the lack of measurement of the most important factors, the example of the theory of heat is most instructive; before the development of the mathematical theory the possibilities of quantitative measurements were less favorable there than they are now in economics. (p. 3)
>
> The historical development of the theory of heat indicates that one must be extremely careful in making negative assertions about any concept with the claim to finality. Even if utilities look very un-

> numerical today, the history of the experience in the theory of heat may repeat itself, and nobody can tell with what ramifications and variations. (p. 17)

We can now say that this assumption is incorrect, because we cannot avoid subjective methods in measuring many variables in economics (e.g., the "aesthetic value of a building" or the "attractiveness of a profession"). All utility theory is based on the possibility of quantitative measurement, but people describe real problems in natural language and not in terms of probabilities and utilities. That is one of the main reasons the axiomatic methods do not work.

I accordingly conclude that the optimality principle cannot provide interesting results for problems in which people have freedom of behavior.

In recent years psychology has invaded economics. We also observe attempts to find neurophysiological explanations of human behavior. The better understanding of human behavior arising from these many fields of research will enrich economics, which has the task of describing, explaining, and predicting human decision-making behavior.

Why optimality is not worth arguing about

Stephen E. G. Lea
Department of Psychology, University of Exeter, Washington Singer Laboratories, Exeter EX4 4QG, England
Electronic mail: *lea@exeter.ac.uk*

Schoemaker argues that the optimality principle is essentially a heuristic: It helps scientists discover valid descriptive statements about the world. Its claim to acceptance therefore depends on its productivity, not on its truth as such. Unfortunately, since Schoemaker rejects both predictive power (as "fundamentally unsatisfying") and causal process (as essentially subjective and "ill-understood," sect. 4), he leaves us without a reliable criterion for assessing how productive a heuristic is. Such iconoclasm is unfortunate, because it risks diluting Schoemaker's most important point, which is that the truth or falsity of optimality is not worth arguing about. This is a very important lesson for those who work at the interface between sciences that are at present dominated by optimality laws (e.g., ecology and economics) and sciences that are not (e.g., psychology and sociology).

Consider, for example, the interface between economics and psychology. Lea et al. (1987, Chapter 5) suggested that arguments about the economic rationality principle (a paradigmatic case of optimality used as explanation) has seriously distorted the development of economic psychology. Too often, research takes the sterile form of psychologists investigating some economic phenomenon, finding that behaviour is irrational, and arguing in consequence that economists' reliance on the rationality principle is misguided.

If optimality is a heuristic, to be judged by the descriptive generalizations it produces, a different inference follows from a disagreement between (psychologically) observed behaviour and (economically) predicted rational behaviour. What we should deduce is that (i) the economic characterization of the phenomenon is wrong or incomplete, *or* (ii) the psychological characterization of the phenomenon is wrong or incomplete, *or* (iii) we have not understood correctly how to aggregate from the behaviour of individuals to the behaviour of the economy as a whole (or, of course, any combination of these). In other words, what we have is a disagreement between descriptive statements, not a disagreement about whether or not people are rational. Hence, if economists take the psychologists' data seriously they may discover deficiencies in their analysis, and in due course replace it with another – which will also be derived using the rationality principle. Equally, if psychologists take economic analysis seriously, they may discover limitations in their data, and in due course supplement them with new data that lead to different generalizations; but the new psychological description is unlikely to be based on optimality principles.

Some, probably most, economists reject this attempt to bypass the argument about rationality. Pen (1988, p. 405) puts it neatly: "Some of us have a general theory of behaviour . . . which is the very core of economics . . . that the world we happen to live in is a neoclassical one." In other words, optimality is a true and universal statement about economic behaviour. Why do economists hold so firmly to this belief? Partly, no doubt, because economics is one of the fields, as Schoemaker argues, where there is a relatively strong prima facie case for optimality.

That case has two bases. The most discussed is the fact that the elements of the economy, human individuals, show rational foresight. But the disparity between the kind of computing power needed to maximise the objective functions of modern economic theory and the known limitations of human thinking means that this argument is no longer defensible. A more plausible basis for optimality is that economic elements, particularly firms, are subject to a process of selection in a competitive market (Hirshleifer 1977). But even natural selection is likely to produce not the best possible solution to any problem but the worst viable solution (Lea 1984, Chapter 1), and there seems to be no reason why competition should do any better.

The real attraction of rationality theory for economists is surely, as Schoemaker argues, its heuristic value. It is a universal tool. Whatever the problem, the principle of rationality can be applied to produce at least a first hypothesis. So long as that hypothesis is then submitted to empirical test, whether at the economic or the psychological level, that is all to the good. Only when generalizations about behaviours are assumed to be valid because they have been derived from rationality does the rationality principle become scientifically harmful. In neither case can we say anything about the truth or fallibility of rationality itself, however. That is indeed a matter that is not worth arguing about.

The example of psychology: Optimism, not optimality

Daniel S. Levine
Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019
Electronic mail: *6344dsl@utarlg.bitnet*

The question whether actual functioning can be considered optimal according to some criterion is of particular importance in the behavioral sciences. Are human-made miseries such as war, income inequality, environmental damage, and widespread ignorance the result of optimal human functioning? The "conscious striving of people and the presence of competition and selection" (target article, sect. 7) incline some people to believe that is so; hence, ironically, an absolute belief in optimality does not lead to *optimistic* conclusions! I believe, rather, that "conscious striving" is only part of the complex interplay of factors determining human or animal decisions. Moreover, a number of recent articles in neural network theory bear on how this interplay may function.

The attempt to place behavior under the rubric of optimality has infiltrated connectionist modeling, though it has not achieved a central place in that field. For example, Klopf (1982) has proposed that both single neurons and brain regions "strive" to "obtain" positive electrical polarization and to "avoid" negative polarization, and that all behavior is explained as maximizing positive polarization of a "controlling" brain area (the reticular formation and midline thalamus). His more recent articles (e.g., Klopf 1988) have not quite followed this line, however, but simply used learning based on such a principle as a component in conditioning and robotic models. Kirkpatrick et al. (1983) (referred to in the target article) developed the principle of simulated annealing, which provides one possible mathematical technique for shifting from a suboptimal to an optimal steady state in a neural system. Simulated annealing, however, has thus far had more influence on the design of nonbiological algorithms to solve applied optimization problems than it has on the design of biological cognitive theories (though I can imagine possible uses for simulated annealing in psychotherapy).

A large number of phenomena in experimental psychology are difficult to explain under an optimality rubric; some of these are reviewed in section 8 of Levine (1983). One example is Gray & Smith's (1969) partial reinforcement acquisition effect (PRAE), in which intermittent reinforcement is preferred to continuous reinforcement, presumably because of the element of surprise. The PRAE owes its existence to the choice of short-term over long-term satisfaction. Another example is the self-punitive behavior that can result if an animal first learns a response to escape shock and then is shocked for making the response (Solomon et al. 1953). Finally, there are the preference reversals found by Tversky and Kahneman (1974; 1981), discussed below.

In those neural network theories that seem to have the most power to explain cognitive data, optimality is only one of many influences on behavior. These models also reflect a strong possibility that several subsystems in the brain are each maximizing different objective functions, but there is no global objective function for the whole brain. This supports the statement of the target article (sect. 2) that "optimality principles may be postulated which collectively do not add up to a coherent whole."

Most important, there are times when an organism is performing some overall function optimally or close to optimally, but its minute behavior is in other respects suboptimal. In areas as diverse as motor control, vision, and decision making under risk, there are network models that recognize this distinction. For example, Bullock and Grossberg (1988) model planned arm movements using a network in which present position is compared continually with a target position, but the intermediate trajectory points are not predetermined. The same researchers contrast their model with a competing one (Flash & Hogan 1984) in which the entire trajectory consists of points at which a variable called "jerk" (the rate of change of acceleration) is minimized. The "minimum jerk" model, Bullock and Grossberg contend, does not allow for performing the same movement at a variable speed or for resuming a movement after it is interrupted.

In the visual area, Levine and Grossberg (1976) and Grossberg and Mingolla (1985) explain various visual illusions as byproducts of an overall system designed to compensate for imperfections in the uptake of stimuli by the retina. Such illusions cannot be considered "optimal behavior" in themselves. Even less optimal are the "cognitive illusions" discussed by Tversky and Kahneman (1974; 1981), whereby preferences in risky situations (involving either money or lives) can run counter to rational utility maximization in characteristic ways. Such preference reversals were modeled by Grossberg and Gutowski (1987) using a neural network architecture called the gated dipole, which had been designed earlier to account for adaptive responses to stimulus changes.

Such nonoptimal behavior is often thought to be incompatible with Darwinian evolutionary theory. I believe that this argument is effectively answered by Gould (1980), who states that evolution should not be equated with progress. In reviewing Darwin's work, Gould (p. 50) says that "organisms are integrated systems and adaptive change in one part can lead to nonadaptive modifications of other features." Also, "an organism built under the influence of selection for a specific role may be able, as a consequence of its structure, to perform many unselected functions as well." While traits compete for survival, at any given moment some traits will be present that are not optimal but not immediately lethal.

Some effects of specific brain damage or particular mental disorders can be seen as exaggerations of suboptimal behavior that occurs in normal organisms. For example, Levine and Prueitt (1989) model various cognitive effects of frontal lobe damage (perseveration in formerly rewarding choices, or excessive attraction to novelty), by constructing neural networks in which the weakening of a specific connection simulates these effects. These networks include separate, interacting subsystems causing tendencies toward attentional bias in favor of previously rewarding events (*affect*); toward continuation of previously rewarding events (*habit*); and toward selective enhancement of responses to novel events (*novelty preference*). Each of these subsystems by itself has obvious survival value. The balance among these subsystems, which is disrupted in frontally damaged individuals, is imperfect even in normals, however. This is why novelty fetishes and perseveration are common in individuals who are not noticeably brain damaged.

The further development of neural network models incorporating affect, habit, and novelty as well as rationality (for a review, see Levine & Leven 1990) is likely to have an impact on optimality theory in the social sciences, particularly economics. Because economic actors are subject to the laws of human behavior, economic theories can ultimately lead to more accurate predictions if they include these nonrational psychological elements. For example, Heiner (1983) argued that economic agents, when making quick decisions on incomplete information, sometimes act in a manner less determined by rational choice than by their own habits. Rational choice theories are also inadequate to explain product preference reversals mediated by changes in motivational context. Using affect and novelty, Leven and Levine (1987) give a neural network explanation for one such reversal: the preference for "new Coke" over "old Coke" in test situations, followed by the failure of "new Coke" in actual buying situations.

In short, I believe that, although optimality is indeed "an organizing principle of nature," it is only one of many such principles and is far from universally valid. In psychology and neuropsychology, there are other organizing principles that cause actual behavior not only to deviate from the optimal, but to deviate in some repeatable ways. Hence, the search for explanatory theories that are elegant and quantifiable need not lead us in directions biased by the "heuristic metastrategies of optimality theory."

The characteristic deviations between prevalent and optimal behavior, however, could make optimality valuable as "a normative principle for individual rationality and social organization." The element of choice in human behavior does not guarantee that optimal choices will be made, but suggests that optimal choices are available. Recent theories of neuromodulation in neural networks (cf. Levine & Leven 1990) hint at possible mechanisms for rapid context-dependent switches between competing patterns of behavior. It remains for network theorists and social scientists, in collaboration, to devise strategies (environmental as well as pharmaceutical) for controlling such switches in desired directions. Thus, relegating optimality to a subsidiary place in social scientific theory actually leads to a deeper optimism about our ability to transcend current behaviors.

Straining the word "optimal"

James E. Mazur
Psychology Department, Southern Connecticut State University, New Haven, CT 06515

As Schoemaker has discussed, the principle of optimization has been popular in many of the social sciences, including economics, psychology, and behavioral ecology. For example, some economists and psychologists claim that people tend to make choices that optimize their "subjective utility" (e.g., Becker 1976). Some behavioral ecologists state that animals tend to make choices that optimize their energy intake per unit time, their ability to mate, or some other variable (e.g., Krebs & Davies 1978). On the other hand, critics of optimization theory have argued that it is not a useful model of either human or animal behavior. I believe that the problems with optimization theory can be illustrated by examining self-control choice situations, which involve a choice between a small, fairly imminent reinforcer and a larger, more delayed reinforcer.

Imagine an experiment with 60 trials a day in which a rat must choose between one food pellet delivered after a 5-sec delay and two food pellets delivered after a 30-sec delay. No matter what choice the rat makes on one trial, the next trial begins 60 sec later. One advantage of this type of experiment is that many of the complicating factors of real-world choices are eliminated. The two choices have identical response requirements (one lever-press), identical total durations (60 sec), and reinforcers of identical quality (food pellets). The only differences between the two alternatives are the number of pellets delivered and when they are delivered. At least on the surface, the predictions of optimization theory may seem unambiguous: To maximize the number of reinforcers received (or the rate of energy intake), the animal should choose the two-pellet option on every trial. In choice situations like this, however, rats (and other animals) do exactly the opposite – they choose the smaller, more immediate reinforcer on almost every trial (e.g., Ainslie 1974; Logan 1965; Mazur 1988).

That animals (and people) will often choose a smaller but more immediate reinforcer has been known for some time. How do advocates of optimization theory deal with this "impulsive" behavior? One strategy is to retreat from the prediction of long-term reinforcement maximization and propose that delayed reinforcers are discounted because their delay and uncertainty makes them less valuable (e.g., Rachlin et al. 1981). But if optimization theory is falsifiable, it should make some specific predictions about the nature of the temporal discounting function. A common assumption in economics (e.g., Becker & Murphy 1988) is that temporal discounting follows an exponential function of the form $V_D = V_O \exp(-KD)$, where V_D is the subjective value of a delayed reinforcer, V_O is the value of the same reinforcer if it were delivered immediately, D is the delay duration, and K is a constant that reflects the discount rate. The exponential equation is a rational discounting function if the costs of waiting for a delayed reinforcer are constant over time. In monetary investments, for example, this equation can be used to calculate the current value of a sum of money to be received at some future date, taking into account the interest that is lost because of the delay. It seems reasonable to assume that, for animals, waiting for a delayed reinforcer involves costs that are analogous to lost interest in monetary investments. Therefore, if animals' choices in self-control choice situations were consistent with an exponential discounting function, they would also be consistent with this modified version of optimization theory.

Unfortunately for optimization theory, there is considerable evidence that neither the choices of animals nor those of people follow an exponential discounting rule (Ainslie 1985; Rachlin & Green 1972). For animals, the data are instead consistent with a hyperbolic equation of the form $V_D = V_O/(1 + KD)$, with all variables defined as before (Mazur 1987). It is not yet known whether this equation is also appropriate for human choices, but it seems fairly certain that the exponential is not (Herrnstein & Mazur 1987).

Suppose it turns out that the hyperbolic equation (or some other one) makes the most accurate predictions for human and nonhuman behavior when delayed reinforcers are involved. It would then behoove advocates of optimization theory to explain how this behavior is consistent with their theory. Perhaps someone could make a case that hyperbolic temporal discounting was indeed the optimal strategy during our collective evolutionary past. Perhaps someone could relate the hyperbolic function to memory decay and claim, using the logic of "constrained optimization," that this is the best we mortals can do, given our imperfect memories.

I have no doubt that some such explanation could be developed. The explanation might even be illuminating and serve as a catalyst for further research. It should be clear, however, that such post hoc explanations make optimization theory unfalsifiable. In addition, calling behavior "optimal" in cases like this gives the misleading impression that people and animals make choices that are rational and optimal in the long run. The many examples of impulsive behavior we see around us (eating too much, drinking too much, smoking, overspending) do not appear to be the behaviors of ideal decision-makers. I would suggest that impulsive behavior reflects neither long-term optimization, nor constrained optimization, but *strained* optimization, because calling this behavior "optimal" strains both the testability and the everyday meaning of the word.

ACKNOWLEDGMENT
Preparation of this commentary was supported by Grant MH 38357 from the National Institute of Mental Health.

Complexity and optimality

Douglas A. Miller and Steven W. Zucker
Computer Vision and Robotics Laboratory, McGill Research Centre for Intelligent Machines, McGill University, Montréal, P.Q., Canada H3A 2A7
Electronic mail: *doug@moe.mcrcim.mcgill.edu*

The ubiquity of optimality, as reviewed by Schoemaker, raises a curious paradox in the theory of computation: Seeking extrema can be enormously expensive. From a computational perspective, then, we can ask: How is it possible that nature is computing all these optima?

On the one occasion when Schoemaker (indirectly) raises the issue, in talking about "intractable combinatorial problems," he does so only to downplay its significance, suggesting that such problems can somehow be "optimized" by "analog models" such as simulated annealing.

This statement leaves the reader with the mistaken impression that such intractable (NP-hard) problems really are tractable if only we write clever enough computer programs. In fact, these problems are unsolvable, barring some stunning breakthrough in mathematics or physics (Garey & Johnson 1979; Vergis et al. 1986) for any practical purpose. Thus no practical theory ought to depend on being able to solve them. In ignoring NP-completeness, Schoemaker loses a major tool for deciding whether optimization is a practical concept in a given situation, and often even what *kind* of optimality is practical (e.g., global or local). In the simulated annealing example, what Kirkpatrick et al. have done is to substitute randomized *local* optimality for global optimality, primarily because the latter problem is NP-complete and the former is not.

To illustrate, consider a problem related to visual processing, that of labeling line drawings of children's blocks in the well known "Blocks world" (Kirousis & Papadimitriou 1988). The task is to decide (label) which block faces are in front of which,

which face intersections are concave and which are convex, and so on, so that all pairs of labels are physically consistent. It has been shown (Kirousis & Papadimitriou 1988) that this problem, for certain very odd but physically possible drawings, is NP-complete.

Computer programs have been written to solve this problem for any instance, but because the problem is NP-complete there are modest-sized instances that would take more time to solve than the age of the planet. Humans are very good at accurately labeling certain line drawings, but NP-completeness says that there should be instances where they either go into a sort of endless loop, or else simply fail to resolve ambiguities.

If we are posing this as an optimization problem (that of minimizing the number of paired inconsistencies in the labels), it seems very unlikely, based on the above analysis, that any biological system would attempt to label this drawing in a way that was guaranteed to be optimal. When presented with an impossibly complex drawing, rather than entering a seemingly endless loop, we just look at it for a while, scratch our heads, and walk away. A more likely approach is some sort of neuronal relaxation that favors quick identification and avoids computer tree searches.

It turns out that an exact mathematical parallel to the labeling problem can be made with economics and the theory of games (Miller & Zucker 1990), for we can express the above NP-complete labeling problems as n-person games in which each player gets a payoff of 1 for each other player with whom his strategy is consistent and zero otherwise. Thus there are games for which there is an optimal global strategy in which all players receive their maximum possible payoff, and yet such a strategy is effectively not computable. [See also Maynard Smith: "Game Theory and the Evolution of Behaviour" *BBS* 7(1)1984; and Caporael et al.: "Selfishness Examined: Cooperation in the Absence of Egoistic Incentives" *BBS* 12(4)1989.]

This implies that in general the problem of finding an *efficient* cooperative solution to an n-person game (in the sense that there is no other solution where everyone does as well and someone does better) is effectively not computable.

So what happens when such a game is played? One kind of solution that *is* effectively computable in a large class of cases where an efficient solution is not (Miller & Zucker 1990) is the Nash equilibrium, where each of the players settles for a distribution of strategies that maximize their own respective return, given that all the other players hold to their Nash solutions. With respect to total payoffs, however, what we have is not global optimality but merely *stationarity.*

Such notions of stationarity and local maxima seem much more plausible than those of global optimality; that is, of seeking the absolute "best." They are widely applicable to modeling the earlier stages of the primate visual system (e.g., Zucker et al. 1989), and avoid many of the pitfalls inherent in global optimization schemes (e.g., Blake & Zisserman 1988; see Tsotsos 1988 for a related complexity analysis). [See also Tsotsos: "Analyzing Vision at the Complexity Level" *BBS* 13(3)1990.] We believe they provide one of the intuitive building blocks for abstracting theories of nature.

Two dynamic criteria for validating claims of optimality

Geoffrey F. Miller

Psychology Department, Stanford University, Stanford, CA 94305
Electronic mail: *geoffrey@psych.stanford.edu*

How can we validate claims of optimality? Perhaps the central issue in Schoemaker's target article is: how can we distinguish real, *inherent* optimality from "postdictive," *apparent* optimality? Schoemaker shows an acute awareness of the human mind's constraints and competencies; those constraints relevant to optimality-attribution may be illuminated by the following evolutionary-psychological argument (in the style of Cosmides & Tooby 1987). Suppose the human mind has evolved to make attributions of optimality to certain teleological systems and agents encountered in its Pleistocene environment of evolutionary adaptedness. For example, perhaps we attributed (e.g., intuitively) some form of (qualitatively) optimal foraging behavior to dangerous carnivores because doing so increased our chances of survival and reproduction. (Assuming adaptive competence in one's competitors and enemies is generally more prudent than assuming incompetence.) More formally, perhaps we evolved the ability to construct mental models of carnivores as human-flesh-intake maximizers, and perhaps this made us better able to position ourselves on certain safer portions of those animals' cost-benefit curves – that is, to make it not worth their while to try to eat us.

Although such evolved optimality-attribution mechanisms may have been well-adapted to certain physical and social situations in the Pleistocene, their extension into abstract scientific domains would be problematic. Perhaps our optimality attributions have been over-generalized. For example, the causal structures of scientific domains may not be isomorphic to those of any domains for which we have evolved effective optimality-attribution mechanisms. Although Schoemaker does not put the issue in quite these terms, he does show a healthy skepticism about projecting optimality into systems where there is none. In particular, he identifies attribution and confirmation biases we may show with respect to our optimality heuristic.

The real issue is: How can we recognize and validate true, inherent optimality? I am afraid that one of Schoemaker's most central propositions will be overlooked: "Each optimality principle . . . begs for an associated process explanation that describes causally, within the constraints of an organism or system, how it operates" (sect. 6.4, para. 2). I would call for two specific sorts of process explanations to distinguish real optimality from other phenomena.

The criterion of dynamic adaptiveness. Humans may tend to anthropomorphize, to attribute what Schoemaker calls "forethought," "assent," "potency" or "consciousness" of self (sect. 4, para. 5) to systems that are not really optimizing, but merely equilibrating. Equilibrating systems move along a trajectory in some state-space until absorbed by some attractor (an equilibrium point or limit cycle). Membrane physics, electrostatics, equilibrium chemistry, annealing, and entropy maximization all seem to involve equilibration rather than optimization. More abstractly, the physical principles of least time and least action could be regarded as results of relaxation in a state-space that includes a temporal dimension. The attractors into which such systems fall may look like the "goals" of those systems, but this seems a rather cheap kind of goal. Although spherical soap bubbles, for example, could be said to be optimizing volume enclosed given surface area available, the teleological connotations of optimization (which Schoemaker views as central to the concept) seem inappropriate for such equilibrating systems.

Real optimization, by contrast, seems to be a phenomenon associated exclusively with what Prigogine and Stengers (1984) call "dissipative systems": systems that work to maintain themselves against entropic breakdown. Thus, claims of systemic optimality should be validated by investigating whether the system's dynamics involve equilibration (collapse to an attractor or limit cycle) or ongoing optimization (which maintains the system "at the edge of chaos," away from attractors – see Langton 1990). I hesitate to invoke something like a Calvinist work ethic for dynamical systems to distinguish whether specific physical systems are "really" optimizing and fighting entropy, or "merely" relaxing in state space, but, as Schoemaker recognizes, the notion of optimality already includes a kind of moral, metaphysical, or evaluative dimension. We may as well recognize that dimension and sharpen our definition of optimality so it excludes processes whose dynamics do not conform to the teleological expectations generated by our scientific terms.

The criterion of historical adaptation. Dissipative systems generally result from some cumulative process of variation, competition, and selection, so, in this view, cumulative selection would be the central process capable of producing real optimality in nature. Maintaining complex systems far away from equilibrium is difficult; it requires very special design features well adapted to the system's environment. Single randomly generated designs are generally far from optimal (i.e., they are unable to maintain the system against entropic gradients), and simple selection operating on a single initial population of designs rarely happens to find anything resembling a global optimum.

Evolutionary biologists attempt to distinguish real biological adaptations from nonadaptations (or concomitants of adaptations) by criteria not only of form and function, but of phylogeny. I would suggest a second, analogous criterion for identifying optimality: consideration of the adaptive process that gave rise to the system in question. Can one identify a specific dynamic process of cumulative selection that could have produced the optimal design or strategy in question? If not, the system may be neither dissipative nor optimal. Identifying the adaptive process responsible for the cumulative selection of an optimal design does not just *validate* the design as more plausibly optimal, but it also *explains* the historical origin and functional nature of the design, and supports our analysis of *what is being optimized.* Thus, the two process questions we need to ask are: What are the system's short-term dynamics as it approaches "goals" (*the criterion of dynamic adaptiveness*), and what are the long-term selective dynamics that gave rise to the system (*the criterion of historical adaptation*)?

Applying the criteria. Optimality arguments in economics strike biologists as weak because economic systems do not typically show cumulative selection of economic entities that is truly analogous to the cumulative natural selection of organisms. Economic systems do show *continual selection* in the sense that individuals or corporations can go bankrupt or succeed, but cumulative selection would require two additional features: (1) *heritable* variation in economic strategy between individuals or corporations (i.e., offspring corporations or individuals actually resemble their parents' economic strategies more than they do the strategies of the population at large); and (2) economic success literally increases the *number* of offspring (i.e., the literal *number* of individuals or corporations produced). Without both of these conditions, there can be no analog of natural selection in the economic realm. Simple continual selection could not suffice to produce optimal economic strategists, any more than the simple survival or death of flatworms without inheritance or reproduction would suffice to produce whales, sharks, or squid after a few millennia.

This is intended not merely to criticise economic theory, but to illustrate the pitfalls of asserting optimality without identifying a specific adaptive process of cumulative selection to validate the historical origins of the optimal design or strategy. Optimality arguments in physics are even more problematic. Even the anthropic cosmological principle advanced by Barrow and Tipler (1986) seems to be a simple continual cosmological selection principle, not a process of cumulative cosmological selection that produces universes "well-adapted" to some unspecified (and unspecifiable?) set of constraints. These two dynamic criteria would also keep sociologists and anthropologists from advancing "functional" explanations of social-level structures without advancing explanations of the historical selection process that led to the generation of the "optimal" social institution, ritual, or practice. I would contend that not only is reference to an adaptive process of cumulative selection required to validate any assertion of optimality but it is also required to validate any assertion of *functionality* per se. (The functionality of human artifacts presumably results from the cumulative selection of alternative designs in the inventor's imagination.)

Arguments about *apparent* optimality could and should be used in any science where they can serve as cognitive shorthand for more complex causal or teleological processes, and their role as a cognitive shorthand is explicitly and carefully recognized. But arguments about *inherent* substantive optimality should be used only in those sciences that deal with adaptive systems subject to processes of differential replication and cumulative selection – that is, biology and its special subfields, psychology and neuroscience. Optimality arguments in other scientific domains can be supported only if their proponents are prepared to identify an *actual* cumulative selection process – or to defend why a cumulative selection process was unnecessary to produce the optimal design or strategy. Reference to processes that bear a superficial or metaphorical resemblance to natural selection is insufficient.

The infinite regress of optimization

Philippe Mongin
Delta, Ecole Normale Supérieure, 48 Boulevard Jordan, 74014 Paris, France
Electronic mail: *corsec@bucllw11.bitnet*

The critique of the optimizing account of individual decision-making has generally emphasized either allegedly empirical failures ("in the real world businessmen do not maximize") or more subtle methodological difficulties of the sort usefully discussed by Schoemaker. This commentary deals with an altogether different class of problems that are best referred to as *logical* ones. As Schoemaker also briefly indicates, there is an infinite regress lurking behind the use of maximizing concepts by decision theory as soon as the latter relaxes the tacit assumption of zero informational and computational costs. Maximizing a constrained objective function is indeed a costly procedure on the agent's part. Supposing that one's search and algorithmic costs can be assessed, these should be made part of one's choice situation. The result is a metaoptimal decision that may or may not upset the optimal one. Supposing that the costs of the procedure used to reach the metaoptimal decision can in their turn be assessed, they should be made part of the agent's choice situation, and so on *ad infinitum.*

There are two ways of understanding the infinite regress just sketched: (a) as a threat to the consistency of the agent's decision criterion, or (b) as a threat to the consistency of the observer's decision theory. It is not clear to what extent agents can assess higher-order costs. A reasonable guess is that more often than not they discover them once they have been incurred; hence there is little sense in engaging in higher-order decision making (except in the case of repetitive decisions). The safe interpretation of the infinite regress is in line with (b) rather than (a): Supposing that the decision theorist knows all the relevant costs, is there a logical level at which his recommendation is *reflectively consistent*, that is, is not upset by the underlying logic of the theory?

Now, the sort of stability property with which the optimizing theory of decision is or is not endowed can be made precise in two ways: (1) There is a logical level n where the $(n + 1)$-optimal decision endorses the n-optimal one; (2) there is a logical level n where the $(n + 1)$-optimal decision endorses the use of optimization at the n-level. That is, the $n + 1$-optimizing decision may be concerned with either n-level pointwise decisions and their costs or n-level decision criteria and their costs. If (1) is violated, the optimizing theorist's recommendation will oscillate in the action space as $n \to \infty$. An example of a violation of (2) is when the theorist shifts back and forth between optimization and satisficing when $n \to \infty$ (the induced recommended action will also no doubt oscillate in this case).

The above discussion should to some extent be familiar to computer scientists, because they obviously have to face the problem of optimal computations and can give a relatively nonarbitrary meaning to the elusive concept of higher-order costs. Strangely enough, economists have not paid due attention to the infinite regress of the optimizing theory of decision, despite the fact that they are the main proponents of this theory. (The few references in the economics literature are Göttinger 1982; Mongin & Walliser 1987; Winter 1975.) This is all the more surprising given that economists have worked out a "theory of search" (Stigler 1961) on the grounds that optimal decisions are, indeed, costly to make.

Discarding objections that could be raised against the shaky variant (a) of the problem, the following seemingly powerful counterargument remains: The infinite regression critique is irrelevant because it affects optimization and alternative theories in exactly the same way. This argument equivocates on the meaning of the critique. It is true that, for example, Simon's (1983) "satisficing" model gives rise to an infinite regress of its own. For any theory T that recommends d, to know that there were decision costs to d and that T is a theory of *rational* decision-making (rather than a nonnormative one) is enough to raise doubts as to whether d should have been recommended after all. Hence the infinite regress *itself* is by no means limited to optimization, but it may or may not converge, depending on the particular theory at hand. The results of the economist's "theory of search" as well as sheer common sense would suggest that convergence in the sense of (1) is more difficult to secure with optimizing than with nonoptimizing theories.

To be specific, Stigler's search model exhibits metaoptimal solutions that are typically different from the optimal solutions. This occurs because the model's monotonicity and regularity assumptions make it worthwhile for the agent to incur a "suboptimality cost" to lower his search cost. The simple trade-off argument can be repeated at any higher logical level. Clearly, it would not apply in the same way, or would not even apply at all, in the context of satisficing. In another example discussed by Mongin and Walliser (1987), a simple assumption connecting the complexity of decision rules with the cost of applying them is enough to destabilize the optimizing theory; that is, the infinite regresses to which it gives rise in this model typically do not converge in the sense of (1).

Don't just sit there, optimise something

J. H. P. Paelinck

Erasmus University Rotterdam, Department of Theoretical Spatial Economics, Postbus 1938 NL – Rotterdam, 3000 DR, The Netherlands

The title of this commentary, taken from an American cartoon showing the professor of economics addressing his students, is illustrative of some methodological excess, at least in my personal field of study.

Schoemaker's target article is provocative, and I would like to comment on it starting from its last section. It is indeed correct to distinguish three types of disciplines: the nonlife sciences, the nonconscious life sciences, and the disciplines of consciousness. It is probably true that at the macrolevel of nonlife enquiries, optimality (extremising under constraints) allows one to derive a good picture of what is going on, though an "explanation" (sect. 6) can be better obtained from the microlevel (sect. 3; Figure 1b). It can be questioned whether optimality is an "exploratory driving force" in the nonconscious life sciences; a blind watchmaker's view (Dawkins 1988) does seem to be a reasonable alternative.

As to the disciplines of consciousness – in particular economics – we would like to present the following comments: Economic theory does seem to be dominated by three ideas: rationality ("narrowly defined as selecting means – the well-known decision variables – that maximize a well defined end": section 5; note that this is a "narrow definition" of the Latin *ratio*!), equilibrium (generally the state of affairs resulting from the solution of the first order optimality conditions) and, implicit in both, optimality, as indicated. Again at the "macrolevel" (not macroeconomics but, say, at a certain level of aggregation), the optimality hypothesis has played a useful part, in allowing one to derive formally specified behavioral equations (e.g., almost ideal demand systems); these are the equations ***in equilibrium***, but to borrow the Prigogine (1980) terminology, economic systems are probably operating "far from equilibrium," and a good way of expressing this is putting down dynamic systems of adaptive equations (maybe better adapted here than in the nonconscious life science in section 7), as we ourselves regularly do in spatial econometrics (Ancot & Paelinck 1983). This also connects with the idea expressed in Paelinck & Vossen (1984, pp. 159ff.) in which optimality is considered a learning process. One should add that "degree of optimality" might differ from individual to individual and from operation to operation: It is probably advisable to be "more optimal" (sect. 2, second last para.) when gambling at the stock exchange than when selecting a summer holiday, which has something to do with the degree of rigor of the constraints and the "forward looking dimension" (sect. 7) that beset the problem. One version of "rational expectations," moreover, allows the integration of that dimension in the adaptive process.

Finally, my view could be expressed as follows:

(1) In practice, economic agents do not "calculate with lightning speed their optimal consumption patterns and output levels" (sect. 1, para. 2), but agents are probably floating all the time at suboptimal levels between extremising and "satisficing."

(2) Extremising principles allow us to derive possible equilibrium values, but sometimes other approaches can be used (conflict analysis, though there, too, solutions are derived by means of multicriteria analysis, again minimising a metric; see van Gastel & Paelinck, in press);

(3) With respect to the last sentence of the "attribution bias" (sect. 6), we think that our minds can probably do no better than use optimal reasoning; other minds (from outer space) would probably be able to do without. Let us mention a trend in this direction, however: Bifurcation, unlike catastrophe theory, does not use a potential to be maximised, though again bifurcations might result from multiple optimal solutions (Paelinck 1990). How hard it is to get rid of optimality. . . .

Optimality as a prescriptive tool

Alexander H. G. Rinnooy Kan

Econometric Institute, Erasmus University, 3000 DR Rotterdam, The Netherlands

Electronic mail: *rinnooykan@hroeur5.bitnet*

Schoemaker's quest for optimality provides fascinating confirmation of the pervasiveness of optimization as an instrument for theory design. It is difficult to conceive of a meaningful answer to Schoemaker's final question whether or not all manifestations of optimality are exclusively in the eye of the beholder, but there can be no doubt about the usefulness of optimization as a recipe for social intervention. The discipline of operations research has made this its hallmark, and it is peculiar that two weaknesses of optimization as a *descriptive* tool transform into strengths in that *prescriptive* setting.

First of all, Schoemaker argues correctly that any descriptive law of science can be mathematically rewritten as the outcome of an optimization process. The point of departure for a prescriptive theory, however, is exactly the opposite one; specification of the optimization model necessarily precedes the possible observation of any empirical outcome. It then turns out that simply in forcing the decision maker to specify his intentions and interests in an unambiguous manner, the optimization approach frequently offers a significant contribution to the quality of the decision process before a single computation has ever been carried out.

Second, people undoubtedly act in a highly irrational manner sometimes, both inside and outside the laboratory. In unflinchingly prescribing optimality however, a behavioural benchmark is provided that cannot be ignored. Hence, especially in an irrational environment, the relentless rigidity of optimization transforms a possibly flawed heuristic of science into a forceful reminder of the standards that should guide our problem solving and decision making behaviour.

Should the quest for optimality worry us?

Nils-Eric Sahlin

Department of Philosophy, Lund University, Lund, Sweden
Electronic mail: *nesahlin@gemini.ldc.lu.se*

The questions I would like to address are: Should the quest for optimality worry us and, if so, what is it that we should be worried about?

In *De Motu,* Berkeley argues that "mathematical things . . . have no stable essence in nature, but depend on the notion of the definer, whence the same thing can be differently explained." This succinct statement comprises the essence of an instrumentalistic view of science. An instrumentalist maintains: (i) that scientific theories lack truth value; they are but computational tools for making predictions, (ii) that theoretical terms such as "electron," "quark" or "utility" have no reference to anything that exists in reality, and (iii) that the functional dependencies expressed by the laws of a theory, for example, "equilibria" or "optima," have (as Berkeley puts it) no stable essence in nature (see e.g., Hacking 1983; Losee 1980; Sahlin 1990).

A scientific realist, on the other hand, rejects the arguments in favour of one or all of these three points. One can, for example, consistently argue that the theoretical entities postulated by a theory exist, but that the functional dependencies expressed by the laws of the theory have no such reference to anything in nature.

For an instrumentalist, the optimality approach would be no more than one of many available tools. With this view of science, equilibria or optima are no more than epiphenomena of our mathematical constructions. Schoemaker mentions a number of inferential biases to which the optimality approach may be prone. Not all of these biases are problematic for someone with an instrumentalistic view of science, however. Instrumentalists do not have to worry about *attribution biases;* they would never claim that since data have been found to fit some optimality model, nature optimizes. Nor must they worry about *illusion of understanding,* because instrumentalists know that theories are simply tools for prediction. Instrumentalists should take seriously what Schoemaker calls *confirmation biases* and *excessive rationalizations,* however. Successful applications of the optimality heuristic in the past can lead to a relentless use of it in the future and an uncritical search for confirmation.

Realists, on the other hand, must worry about most of these biases, depending on what type of realist they are, and they must especially ask themselves whether it is reasonable to assume that nature optimizes.

Schoemaker seems to dismiss parts of this important difference by arguing that the "positivist view that only prediction matters is fundamentally unsatisfying, and optimality principles consequently suffer from being too paramorphic." The fact that our ontological commitments may lead to rather different inferential biases, however (which shouldn't come as a surprise), must be of vital relevance to the scientist. This is an insight that has little to do with the view that only prediction matters.

One way to minimize the risk of unwanted inferential biases, therefore, is to scrutinize our ontological assumptions thoroughly. In psychology and economics, much of the work on human decision making is, as Schoemaker correctly points out, formally expressed in optimization terms. Purely normative theories of decision tell us that it is "rational" to optimize. It is unreasonable that so much research effort has been spent testing these normative theories' descriptive and explanatory validity. It is obvious that these theories do not provide us with a sound theoretical foundation for understanding human decision making. A sound and developed theory of belief and value is missing. Our theory and understanding of human decision making will be rather different depending on whether we adopt, for example, a mentalistic or a dispositional theory of belief, or we argue that a belief is a mental state (i.e., as Frank Ramsey [1929] puts it, "a belief . . . is a map of neighbouring space by which we steer").

An example will show what I have in mind. It is a central idea of the Bayesian paradigm that a decision maker's state of belief can be represented by a unique probability measure. Thus, experiments have been designed to test how good people are as decision makers – a test for optimality. This seems, at least to me, to be a far from fruitful approach. To make it successful we must first say what a belief is. But if a belief is, for example, a mental state, purely theoretical considerations tell us that it cannot be represented by a unique probability measure, at least not if we want to mirror the content of the belief accurately. No one would argue that a reasonably detailed map can be represented by a unique function (of a frugal number of variables). An emphasis on experimental studies, without thorough theoretical foundations, therefore, will generate various inferential biases and results that give us an inadequate understanding of human cognition and decision making.

There is another distinction that has to be considered in discussing the quest for optimality. In most cases of interest, one and the same phenomenon can be mathematically characterized in a number of different ways. It may happen that some of these descriptions make use of a concept of optimality, but not necessarily all of them. A realist would argue that these are, *de dicto,* equivalent descriptions – they have the same extension. They do not have to be equivalent *de re,* however, that is, not all of them need to capture the essence of the studied phenomenon. What might seem to be a characteristic of an investigated phenomenon, thus, may in effect simply be an epiphenomenon of our mathematical description. Bengt Hansson (1988) has shown that confusing *de dicto* with *de re* explanations has led to misunderstandings in the interpretation of modern utility theory. If people act in accordance with the axioms of utility theory, they will, from the point of view of the bystander, act as if they had a utility function and maximized expected utility. Agents do not need to have a utility function, however, nor do they need to be maximizing expected utility.

What this shows is that the search for optimality does not in itself need to be worrisome; it is when it is based on an ill-conceived theoretical foundation that it becomes problematic. My thesis is that the quest for optimality is not as serious a problem as Schoemaker tends to make it – provided that we thoroughly scrutinize our ontological assumptions.

Rational agents, real people and the quest for optimality

Eldar Shafir
Department of Psychology, Princeton University, Princeton, NJ 08544
Electronic mail: *eldar@clarity.princeton.edu*

Delving into fields of enquiry as diverse as economics, physics, chemistry, and biology, Schoemaker raises interesting questions about the role of optimality in scientific theorizing. Optimality may have different roles in different theories. As Schoemaker points out, the appeal of optimality arguments stems in part from our desire for simplicity and elegance in science. But whereas certain simple and elegant theories may be successful, or even right, others are likely to be inappropriate and wrong. In one domain, optimality may look like an organizing principle of nature, whereas in another it may appear to be a naive idealization. Light travelling through water may behave as if it is optimizing (whatever the philosophical interpretation), whereas economic man choosing between alternatives may not. The target article compels us to consider not just the nature of optimality per se, but its relevance and applicability to theorizing in several domains.

Perhaps the best known invocation of optimality principles in the social sciences is the rationality assumption that underlies microeconomic theory. According to this assumption, decision makers behave in a manner that conforms to some very simple principles of rational choice such as the principle of dominance: If one option is better than another in one state of the world, and at least as good in all other states, then that option should be chosen. Another principle is that of invariance: Different representations of the same choice problem, or different methods of eliciting a choice, should yield the same preferences (Tversky & Kahneman 1986). What is the status of these and other rational-choice principles? Do they constitute the core of human economic behavior? Are they, rather, part of an unrealistic view of human nature? The answer to these questions depends largely on whom we take the decision makers to be. If we are thinking of von Neumann & Morgenstern's (1947) hypothetical rational agents, who populate the world of many economists, game theorists, mathematicians, and statisticians, then these optimality desiderata are organizing principles or, better yet, defining characteristics. On the other hand, if what we have in mind are *real* people, like us and our friends, then these same optimality principles are nothing more than inapplicable assumptions, wishful ideals.

Schoemaker suggests that "optimality in science is a powerful heuristic for describing existing phenomena as well as predicting new ones." To be a successful heuristic, however, optimality should make predictions in the same domain that it describes. Instead, one frequently encounters economic predictions regarding real people, based on descriptions that apply only to rational agents. It is when we predict one population based on the description of another, very different, population that the nature of the optimality heuristic becomes particularly enigmatic.

Although normative theories are characterized by elegant and mathematically sophisticated optimality principles, a rich body of work shows that these theories are not reconcilable with the way real people behave (for reviews, see Hogarth & Reder 1986; Slovic et al. 1989; Tversky & Kahneman 1986). As a consequence of this tension, several theories that retain some of the more normatively appealing principles while relaxing others have been developed in recent years (for reviews, see Camerer 1990; Machina 1982; Schoemaker 1982). According to these theories, decision makers violate certain principles (e.g., independence and transitivity) but not others (e.g., invariance). But who are these theories about? They are not about rational agents, because *they* behave in conformity with *all* the principles and have no good reason to give up any. They are also not about real people, because we regularly violate all the principles, including some of the normatively indispensable ones such as dominance and invariance. There is, it appears, an interrelationship between our criteria of optimality, and what these optimality criteria refer to in a given domain.

The question of whom (or what) science's optimality principles refer to seems pivotal in contemplating what these optimality principles mean. In the theory of individual decision making, this question may lead to remarkably diverse answers. The economic rationality assumption may be considered with reference to at least three distinct groups: rational agents who satisfy all the underlying principles of the rationality assumption; people like us, who perhaps believe in these principles but do not behave as if we do; and other, recently evolved creatures who satisfy some principles and violate others. Optimality (defined in terms of the rational principles of choice) is a law of nature for the first group, an impossible dream for the second, and a compromising *raison d'être* for the third. In reflecting on what optimality is, we need to consider whom it is about. The answer to the question, "What is optimality?" will depend, at least to some degree, on who we think is optimizing.

Extremum descriptions, process laws and minimality heuristics

Elliott Sober
Philosophy Department, University of Wisconsin, Madison, WI 53706
Electronic mail: *esober@wiscmacc.bitnet; esober@vms.macc.wis.edu*

The examples and concepts that Schoemaker cites are rather heterogeneous. Some distinctions need to be drawn. Consider first the example of Fermat's law. It states a *minimum principle*. The law does not say whether it is good or bad for light to follow the path of least time; it just says that light does this. An optimality thesis involves not just an ordering of options, but a value judgment about them. So let us begin by distinguishing minimality from optimality.

The concept of minimality can play a variety of roles. First, there are statements to the effect that the value of some parameter is as low (or as high) as it can get in systems of a particular kind. I will call such hypotheses *extremum state descriptions*. Second, there are statements to the effect that a given kind of system engages in a process of minimizing (or maximizing) some quantity. These diachronic claims hypothesize an *optimizing process*. Third, there are *methodological recommendations* to the effect that one should prefer one hypothesis over another on the ground that the first is favored by some optimality (or minimality) criterion.

Fermat's law says that the path of a light ray moving from one medium to another has a given *minimum property;* it does not say that light engages in an *optimizing process* in which light manages to successively reduce its travel time. Still less does Fermat's law state a *methodological recommendation,* either about how to theorize about the physics of light or about any other subject. So Fermat's law falls into the first of the above three categories, but not into the other two.

Process laws sometimes postulate a monotonic increase (or nondecrease) in some quantity. Entropy goes up in closed systems, according to the strict second law of thermodynamics. Fitness increases in systems that obey Fisher's fundamental theorem of natural selection. Neither of these laws requires that any real system actually exhibit the maximal degree of the quantity involved; each simply says that systems are on their way toward that maximum, whether or not it will ever be attained.

The idea of optimality (or minimality) as a "positive heuristic" in science, which is Schoemaker's main subject, concerns the justification for accepting or preferring hypotheses. Heuristics are methods, not the items obtained through the use of those methods. The content of a statement must not be confused with the methods by which the statement is accepted or favored.

This point is important, for there is no necessary connection between extremum state descriptions or optimizing process laws on the one hand and, on the other hand, the heuristic principles that may tell us to favor such descriptions or laws. Fisher's theorem (understood appropriately as a conditional statement) is a *theorem;* it is a deductive consequence of the appropriate starting assumptions. One does not have to accept an optimality heuristic to believe this law; it suffices to follow the mathematical argument. So Fisher's theorem belongs to the second category mentioned above, but not to the first or third.

Just as an extremum state description or an optimizing process law can sometimes be justified without resort to an "optimality heuristic," so an optimality (or minimality) heuristic can justify a hypothesis that does not itself assert either an extremum state description or an optimizing process law. For example, the method of cladistic parsimony used to reconstruct phylogenetic relationships favors the genealogical hypothesis that requires the fewest changes in character state; but this favored hypothesis does not assert that the number of changes is no higher than this minimum. One must not confuse minimizing assumptions with assuming minimality (Farris 1983; Sober 1988).

Schoemaker asks, "Who is optimizing: the scientist or nature?" The above threefold taxonomy can be used to clarify this question. Extremum state descriptions and laws that postulate an optimizing process make claims about nature. When they are well justified, we may conclude that various natural systems occupy extremum states or engage in a process of monotonic increase (or decrease). What is less straightforward is the third category: When a methodological principle tells us to favor hypotheses in accordance with some optimality criterion, what, if anything, does the use of this criterion presuppose about nature?

Here we must be careful to consider various methodological recommendations one at a time. In standard regression analysis, the best regression line is the one that minimizes the residual variance. This methodology can hardly be said to assume that variance is minimal in the world at large. Regardless of how much variance there is in a given inference problem (and even if the idea of "the amount of variance in nature at large" is not well-defined), there is a well-understood rationale for why one regression line is to be favored over another (Farris 1983; Sober 1988).

Other methodological recommendations are less well understood. Philosophers have puzzled for a long time over the credentials of Occam's razor. Why should we use parsimony or simplicity as a criterion? These minimum principles apparently are used to guide us in what we believe or think plausible. So the mere fact that simpler hypotheses are easier to understand and manipulate, or more beautiful to behold, does not seem enough. The perennial problem has been why we should regard simplicity and parsimony as signs of truth.

Newton laid down as his first rule of reasoning in philosophy that "Nature does nothing in vain . . . for Nature is pleased with simplicity and affects not the pomp of superfluous causes." Leibniz hypothesized that the actual world obeys simple laws because God's taste for simplicity influenced his decision about which world to actualize.

Epistemology since Hume and Kant has drawn back from this way of understanding methodology. The view has taken hold that a preference for simple and parsimonious hypotheses is *purely methodological;* it is constitutive of the attitude we call "scientific" and makes no substantive assumption about the way the world is.

A variety of otherwise diverse philosophers of science have attempted, in different ways, to flesh out this position. Two examples must suffice here; see Hesse (1969) for summaries of other proposals. Popper (1959) held that scientists should prefer highly falsifiable (improbable) theories; he tried to show that simpler theories are more falsifiable. Quine (1966), in contrast, saw a virtue in theories that are highly probable; he argued for a general connection between simplicity and high probability.

Note that both these proposals are *global.* They attempt to explain why simplicity should be part of the scientific method in a way that spans all scientific subject matters. No assumption about the details of any particular scientific problem serves as a premise in Popper's or Quine's arguments. In this respect, their positions are continuous with the tradition stemming from Newton and Leibniz.

In view of the various inadequacies that attach to these global proposals, it is perhaps time to take a more local approach. Let me suggest here a strategy for understanding how simplicity and parsimony criteria function in science, one that I have attempted to implement elsewhere (Sober 1988; forthcoming). When the choice of a hypothesis is justified by appeal to parsimony or simplicity, this must be because some substantive assumption about the world is in play; in practice, Occam's razor is not "purely methodological." But different appeals to simplicity and parsimony make different assumptions in the settings of different inference problems. There is no global principle of parsimony whose justification is subject matter independent.

At the end of his article, Schoemaker lists a set of choices. Is optimality "(1) an organizing principle of nature, (2) a set of philosophically unrelated techniques of science, (3) a normative principle for individual rationality and social organization, or (4) a metaphysical way of looking at the world?" I would suggest that the phenomena Schoemaker considers are so heterogeneous that the answer must be: *all of the above.* Doubtless the concept sometimes plays each of these roles. If this is right, then the way to improve our understanding of optimality concepts in science is to look at concrete examples in some detail, without assuming in advance that there is a single analysis that must work across all the cases.

Avoid the push-pull dilemma in explanation

Kenneth M. Steele

Department of Psychology, Mars Hill College, Mars Hill, NC 28754
Electronic mail: *kms@ecsvax.bitnet; kms@ecsvax.uncecs.edu*

Schoemaker asks about the types of acceptable scientific explanations of "why x occurred." His answer is that there are two types, causal and teleological. In this commentary, I want to make the following points. First, Schoemaker's distinction between causal and teleological explanations is tenuous. It has already been violated in past psychological theory. Second, his distinction misses an important difference. Optimality analyses are often associated with another type of scientific explanation, one that eschews concern over the mechanical order of events. Such analyses are not concerned whether a previous event is pushing or a future event is pulling. Finally, optimality analysis is a tool: Like all tools, it can be used correctly or it can be abused.

What is wrong with Schoemaker's distinction? Consider the following case from psychology: A rat, deprived of food for a suitable amount of time, is placed in a straight alleyway that contains a morsel of food at the other end. The rat, after several such experiences, behaves quite differently from the way it did on its first trip. When it runs, it runs quickly and with the appearance of "going somewhere." What is a good explanation for this rat's new performance? Schoemaker suggests that there are two types of explanations, causal and teleological, which are fundamentally different. The causal one consists of a chain of antecedent "pushes" and effects. In this case, the pushes would probably be described physiologically. The teleological explanation focuses on a goal, and an intention to obtain it. The "pull" from the food-morsel on the rat might be expressed using a cognitive phrase like "food expectancy."

Schoemaker's alternatives makes the analysis by Clark Hull (1930; 1931) of interest here. Hull assumed that the world in the alleyway, for the moving rat, was a sequential flux of conditions (stimuli) and reactions. The changing outer world produced a parallel series of changes in the rat, that is, representation. The rat was "pushed" into action by kinesthetic stimulus aspects of the "hunger drive," which constituted a *persistent* core of internal stimulation (as opposed to the flux of the outer world). With repeated experiences of finding the morsel at the end, such drive stimuli called forth fractional anticipatory goal responses (e.g., the licking of one's lips). These fractional anticipatory goal responses also had kinesthetic stimulus qualities that served to elicit responses. Because the hunger drive was a persistent cause throughout the sequence, the control by *anticipatory* goal responses was also *persistent* throughout. In other words, the fractional anticipatory goal response corresponded quite literally to the "idea" of obtaining food; and an idea or goal was a causal event!

There are two points in Hull's analysis that are relevant to Schoemaker's distinction. The first is that pull and push are inextricably intertwined in Hull's moment-by-moment analysis of the rat's behavior. This is no happenstance. At that point in the history of psychology a similar choice was offered between push versus pull explanations. Hull's model swallowed both and shows that Schoemaker's distinction is not a dilemma. Second, with regard to Schoemaker's emphasis on the metaphysical nature of teleological explanations, the nature of stimulus aspects of a drive-construct and the nature of stimulus aspects of an anticipatory-goal-reaction-construct seem equally impalpable.

The further developments in Hull's theory exemplified his vision of an adequate scientific explanation. Hull's later explanations (1943; 1952) of the rat in the alleyway became more complex, not less, with the introduction of such additional constructs as stimulus intensity dynamism, incentive motivation, reactive inhibition, reaction threshold, and reaction potential oscillation. But all contributed to Hull's vision of an adequate scientific explanation, which was to give a complete moment-by-moment explanation of the rat's behavior.

And that is where optimality analyses are different. The difference can be seen in Schoemaker's "trivial example," which begins with the assumption of the empirical law: $y = f(x)$. Russell (1959) pointed out the advantages of such means of expression. Consider Ohm's law ($E = IR$). There is nothing easily distinguished as *the* cause or *the* effect in the expression relating voltage to current and resistance. It is just as meaningful to make any of the three variables the effect or goal. One could specify that a particular voltage level must be held in the circuit, or that current load was to be minimized, or both.

An optimality analysis is based on an objective function, which mathematically specifies the relationship between some number of variables in terms of such a common dimension or currency as utility or reproductive fitness. It is a tool by which a scientist may specify mathematically assumptions or conclusions concerning a relationship; the formulation can then be evaluated. That the goal is expressed as the maximization or minimization of some variable is more properly understood to mean that there is a systematic preference for or against the goal variable. The extremum aspect of the formulation is akin to that of the operational definition, specifying exactly what is included or excluded.

Optimality analyses are often balance or regulation models. They are attempts to express relationships in a fashion similar to Ohm's law. One positive feature of such models is that they do not immediately demand specification of moment-to-moment causal sequences. There is no concern about filling every moment of time with explanatory entities. They allow us to avoid sterile push versus pull or cause versus effect arguments. Such a difference in the use of models can be seen in comparing the theories of Staddon (1983) and Hull (1943), who have both combined biology and mathematics in psychology. Forty years have made a difference. It is certainly true that the technique may be misused, and one type of misuse is overuse (Gould & Lewontin 1979). But the point is, this is true of all scientific tools and can't be used to decide the worth of an analysis.

Optimal confusion

Stephanie Stolarz-Fantino and Edmund Fantino
Psychology Department, University of California, San Diego, La Jolla, CA 92093
Electronic mail: *p528@sdcc12.ucsd.edu*

The author of this illuminating and thought-provoking target article invites us to decide what optimality "really is," offering four possibilities, as well as "something else still." Although we favor "all of the above," as behavioral scientists we are most interested in the third meaning of the "optimality heuristic": "a normative principle for individual rationality and social organization." Research in one of our laboratories has used optimal foraging theory to provide useful predictions of choice behavior that have been confirmed empirically (e.g., Fantino & Abarca 1985; Fantino & Preston 1988a;1988b). As pointed out in this journal, delay-reduction theory (DRT), developed over the past 21 years in the operant laboratory, generally makes the same predictions as – and can be shown to be equivalent to – the optimal diet model of classic optimal foraging theory (Fantino & Abarca 1985). These theories, however, do not always require behavior that maximizes reward over anything but brief temporal intervals. And when the predictions of these more molecular models, such as DRT, are pitted against a more molar optimality approach, ongoing work in our laboratory (with Wendy A. Williams) shows that choice follows the predictions of DRT and not optimality. Generally, however, behavior obeying the DRT is also optimal at a more molar level, and one is reminded of Simon's theory of bounded rationality, according to which, as Schoemaker notes, "people satisfice rather than optimize." But whereas it is no simple matter to show that hungry pigeons deviate from maximizing food intake (consistent with a more molecular, "satisficing" rule), it is apparently easy to show that the inferential reasoning of human subjects may be dramatically flawed. We offer an example and then attempt to reconcile it with a "satisficing" viewpoint.

The "conjunction effect," reported by Tversky and Kahneman (1983), is an example of base-rate error in human reasoning in which subjects report the conjunction of two events to be more likely to occur than either of the events alone. This effect is quite robust – in fact, our own attempts to eliminate it by modifying our instructions to make the underlying logic of the task more obvious resulted in an even larger effect than before. Since behavior is, we believe, largely a function of past experience, it must be the case that experience does not always push organisms toward greater rationality – at least in the short term.

As discussed by Nisbett and Ross (1980), most of people's inferential errors have no serious consequence in daily life, and may even be inadvertently reinforced. Organisms that are primarily sensitive to short-term (molecular) events, in most cases do adequately in the long-term (i.e., in the molar sense) as well. But optimal or not, behavior cannot be described, explained, or predicted fully without taking into account events on a molecular level.

ACKNOWLEDGMENT
Preparation of this commentary was supported by NIMH Grant MH-20752 to the University of California at San Diego.

The human being as a bumbling optimalist: A psychologist's viewpoint

Masanao Toda
Department of Cognitive Science, Chukyo University, Toyota-shi, 470-03, Japan
Electronic mail: *toda@sccs.chukyo-u.ac.jp*

I agree basically with Schoemaker that "to a large extent, optimality is in the eye of the beholder," with one exception: The beholder, a human being, is in my opinion a true optimizer in his *intentions*. Being a psychologist, I will concentrate in this commentary on elaborating this particular viewpoint.

Let me choose physicists as my first target. It certainly goes without saying that major features of the models of the physical universe that they have built are not those of the optimizing type, despite the existence of a few of such interesting optimizing physical laws as Fermat's principle. (Many of the simple extremum laws of physics are of the trivial kind, however.) What interests me most about physicists is the sharp contrast between this lack of optimality principles in their models and their extremely "optimalistic" attitude. As cited by Schoemaker, physicists seem to be vigorously pursuing their shared aesthetic goal of obtaining the ultimate model of the physical universe, one that is simple, elegant, and symmetric, as well as valid. So physicists are undoubtedly optimizers in their intentions. Note at this point that this contrast confirms the obvious: Having optimalistic intentions has nothing to do with whether or not the models scientists create are also optimalistic.

A more interesting issue here is the relationship between physicists' intentions and their behavior. As intentional optimizers, do physicists try to maximize the aesthetic value of their models by directly manipulating aesthetic dimensions? Occasionally, such as when they try to replace clumsy-looking equations with equivalent but more elegant ones, or when they strive to discover magnetic monopoles to eliminate nonsymmetry between electricity and magnetism. They are not directly optimizing their models in doing so, however, because there is no need. For most physicists the optimal model of the universe already exists, it is merely hidden. So what they are supposed to do is find it; to that end, manipulating aesthetic dimensions might help as a heuristic strategy, because the missing model is already known to be *beautiful beyond comparison*.

Now let me turn to such other disciplines as psychology and the social sciences; they should reveal another aspect of the relationship between scientists' intentions and their behavior. If all people, physicists or otherwise, are intentional optimizers as I contend, then why do psychologists not endorse more strongly optimizing models of human behavior? The reason seems to rest less on the complications of trade-off relations among each person's, say, wealth-optimizing, happiness-optimizing, and similar intentions; it has more to do with people being bumbling optimizers in their *behavior*, even if their intention to optimize remains stable.

An individual person lives in an opaque world, consisting of very many unknown variables, apparently too many for a person to be capable of handling in any straightforward optimal way. Within that opaqueness, or limited visibility, one can do little more to optimize than to choose a better-looking alternative among those presented on each occasion. This type of behavior apparently offers little chance for an optimizing behavior model to succeed, because a person may easily be trapped on top of a local hill (its altitude measured in terms of a personal value scale), or, while trying to climb a steep value-cliff, he may slip and fall into a river and be carried farther downstream. If individuals should behave this way, the preoccupation of psychologists with *causal* models of behavior would appear natural. The controversial point concerns whether or not psychologists need optimizing behavioral models in addition to causal ones.

To consider this point we should first clarify a little what causal models are really for. To make what should be a very long argument short, let me tentatively characterize causal models, either those of a scientist or those of a layperson, as the consequence of an attempt to describe the behavior of some target agent (animate or inanimate) in terms of a set of carefully chosen antecedent conditions. Though there are apparently infinite degrees of freedom for what set of antecedent conditions to choose as the cause of some effect, some choices are practically not considered; for example, to choose as the cause for a glass being broken the event that occurred just a moment before, namely, that the glass was about to hit the floor with a certain velocity, would obviously be useless. So a certain time lag between cause and effect is important: but there are also other elements determining preferences for *hypothetical* causes (their truth values are not my concern here). The characteristics of the preferences may be summarized as follows: Knowing about the actual occurrence (or nonoccurrence) of one of the causes must confer the power to increase one's chance for optimization, for example, through enabling one to flee from a disaster (an effect) before it happens, or through preventing the disaster itself by controlling its cause in advance. So, after all, even causal models are human inventions *caused* by people's intention to optimize. Both causal and optimizing models are therefore needed to establish a satisfactory science of human beings, even though the interrelationships between these two types of models could be quite complicated.

Now let me briefly consider the social sciences, where the target universe is a world full of bumbling optimizers, yet it is still doubtful that society at large has even a semblance of the intention to optimize. Note, however, that these sciences have again been created from human beings' desire to expedite their optimization. People are apparently not satisfied with just having causal guidebooks for coping with local issues, but also aspire to have a sort of global map, however crude, to aid them in making longer-range plans or such collective decisions as policy making. In any case, the real issue for social scientists is how to handle the enormous number of variables they are beset with. There seem to exist two major ways for cutting this Gordian knot: The first is to make simplifying assumptions, as usual. Assume that everyone is an optimizer even in behavior and not just in intentions, endowed with exactly the same dispositions and preferences. Such an assumption might produce some approximate model of human collective behavior for the reasons cited by Schoemaker. The second alternative suggests the use of control theoretic models that superficially resemble true optimization models. Once a certain state of society is declared desirable, one need pay attention only to the *deviation* of the actual state from the desired one, and the efforts of policy makers can be concentrated on reducing it. This should in turn be relatively easy because, if the deviation is *small*, many of the troublesome variables can be rightfully ignored as making only insignificant contributions to the deviations.

Now, let me come to my final remark. Obviously, human beings as intentional optimizers are an outcome of evolution. Causally speaking, Nature does nothing but eliminate relative misfits who have failed to discover a safe haven. Could there be any effective means of traversing this seemingly great distance between a relatively mild causal rule and the creation of something as freakish as an intentional optimizer? As the space for this commentary runs out, I happily leave this tantalizing question open.

Is economics still immersed in the old concepts of the Enlightenment era?

Andrzej P. Wierzbicki
Institute of Automatic Control, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland

Through most of my professional life I have been working on optimization theory, its techniques and their applications in various fields. Schoemaker's target article has convinced me that I should goad economists – at the risk of offending some friends – into accepting a less traditional view of the world, for economists wield optimization like a samurai his sword: The sword is beautiful and they use it ceremonially and ritualistically, but not sparingly or purposefully. This may be beautiful to watch, but have you ever heard of a Japanese team winning the Olympics in fencing?

Schoemaker might have had similar purposes, but I cannot agree with any of his four suggestions about the meaning of optimality. It is the fifth possibility – something quite different.

What is optimality and optimization? Optimality is just a basic concept in optimization theory, which is in turn a part of mathematics. And mathematics is a language – a tool for stating our thoughts more precisely and in shorthand notation, with powerful means for checking the internal consistency of such statements. Mathematicians spend their lives developing and polishing this tool, guided by a specific instinct and taste. They know that the number of theorems that can be proven is infinite and the real question is which to prove next, not how many to prove (in contrast to some economic theorists).

There are parts of mathematics that have integrative character; in a short statement they say more than is apparent at first glance. Optimization theory belongs to them, but other parts might be even more integrative. For example, the concept of separating two sets by a (level set of a) function is also used in optimization theory; this is a highly integrative concept. To say that a function separates, at a given point of a set, this set and a cone shifted to this point, is equivalent to three more detailed statements: that this point is efficient (pareto-optimal) with respect to the ordering induced by the cone; that this point maximizes the function on this set; and that this function is (at least locally) monotone with respect to the ordering. Thus, arguments based on separation of sets simplify proofs and give new insights. Mathematics is full of such integrative statements. Optimality can also be used to state other mathematical concepts. For example, the concept of a subdifferential extends the classical calculus of derivatives; this concept was developed as part of optimization theory, however, and is often stated in its terms. The concept of a projection, of approximation and many others involve optimality.

Being integrative, optimization theory is a very good tool for applied research: Its application in a field of science can give unexpected insights. Such insights must be checked for applicability, but they often prove to be good enough (in a given context and scope). Schoemaker and many economists choose to call this fact the teleological nature of optimality; I would give a different explanation: It indicates instead that the instinct and taste of mathematicians developing optimization theory were correct; optimization provides a truly useful tool for applications – particularly in fields where conservation principles or accounting identities (which can usually be restated in terms of optimality) are important, as in physics or economics. Such fields use optimization extensively, but not necessarily in the most sophisticated way; for example, applications of nondifferentiable optimization started first in other sciences.

We could, however, choose to restate optimality principles (as indicated above) in terms of the separation of sets in an applied field. What would the teleological nature of such a statement be then? That light "travels in order to separate two sets"? And if we used other integrative parts of mathematical language that might provide other insights? Teleology is in the eye of the beholder.

The role of changing basic concepts about our perception of reality. Many researchers, reflecting on the current state and the future of human civilization, conclude that the basic concepts of perceiving time and space, cause/effect, and reality in general have changed considerably during the twentieth century and that this change has played an important role in the transformation of current industrial civilization into a new (postindustrial? information? informed, humanistic, and global?) stage. Some argue that the old concepts reflect a mechanistic perception of reality developed during the Enlightenment era and adopted by industrial civilization.

The industrial civilization concentrated on material resources that are limited – which results, loosely speaking, in a constant-sum economic game. Information resources, when shared, usually increase – and the sum of the game is not constant. Modern biology has largely abandoned the Darwinian (I agree with Schoemaker's Note 6 that it is actually Spencerian) mechanistic concept of the strictly competitive evolution of the "survival of the fittest" and included many other evolutionary principles. On the intersection of biology and decision sciences, new concepts of evolution of cooperation are investigated (I wonder why Schoemaker did not cite the results of Axelrod (1984) and Rapoport et al. (1976) and how he would classify them). In economics, the focus is usually on the "survival of the fittest" – even if more advanced evolutionary concepts are investigated by a few more radical thinkers (Simon, cited only for his earlier works (1957; 1981), or Nelson and Winter (1982), neither quoted). For the "economic man" should be mechanistic: selfish, consistent, and of unchanging taste; how should we then characterize an "informational man"? [Cf. Caporael et al.: "Selfishness Examined: Cooperation in the Absence of Egoistic Incentives" *BBS* 12(4)1989.]

I agree with Schoemaker that optimization, as it is used in most parts of economic theory, leads to a counterposition of mechanical causality and teleology. Meanwhile, however, mechanical causality has been rendered obsolete by several concepts of such modern systems theory as dynamic processes with feedbacks, self-organization and synergy in complex systems, chaotic processes, as well as the more relativistic distinction between a law of nature and a model. The concept of feedback, for example, was developed more than 50 years ago in telecommunication and automatic control as a strictly causal (though more complicated) mechanism. It spread to other fields, however, and made obsolete the mechanical causality in the analysis of more complex systems – to such an extent that I have heard long discussions about whether the concept of feedback is causal or teleological. In economics, even if dynamic processes are investigated and the concept of feedback is used, the mathematical content of a typical academic curriculum is focused on static calculus and optimization, sometimes extending to a few dynamic aspects of the latter, not dynamic disequilibrium processes with feedback. Thus, most of the economic profession is not equipped to investigate dynamic evolutionary processes.

Relativistic physics has prompted other sciences to adopt a more relativistic attitude (in the popular, not the physical sense of this word) toward their own results, which – even if stated in mathematical language – can consistently be described only as models of reality, not laws of nature. In a model, however, teleology is always "as if," often a kind of anthropomorphism – although a good model should help in understanding (which is more than explaining) the reality. Models of the "black box" type can be useful only in a very ad hoc sense – hence, Friedman's (1953) arguments are indeed weak. Moreover, utility maximization heuristics are often taught to students as a basic law of the economic world – whereas they obviously only comprise a model of rather limited applicability.

I use the term"utility maximization heuristics" here, not "rationality heuristics" as Schoemaker does, because it has been argued elsewhere (e.g., Sopron 1984) that utility maximization is only one of many possible frameworks for perceiving rationality. For example, some Japanese economists argue that utility maximization heuristics were developed in American economics because they reflect the individualist culture that was historically useful for conquering a large, almost empty continent, whereas the more collectivist culture of Japan would require harmony maximization heuristics – and harmony is not another type of utility, because it describes a group, not an individual. The best evidence that other cultures might perceive rationality differently is the fact that the long list of references in Schoemaker's target article contains almost exclusively American and British sources. I could add many others reflecting different viewpoints, but even listing those references to which I owe much of the thoughts presented above would exceed the limits of this commentary.

I should add, however, that the question in the title of this commentary is emphatically not restricted to American economics; it certainly applies even more (and has been posed), slightly differently, by Jozef Pajestka) to economics in my own country.

Author's Response

The strategy of optimality revisited

Paul J. H. Schoemaker
Center for Decision Research, Graduate School of Business, University of Chicago, Chicago, IL 60637
Electronic mail: *fac_paul@gsbacd.uchicago.edu*

Not surprisingly, my wide ranging target article on optimality elicited a broad spectrum of reactions. Of the 27 commentaries, 15 came from psychology, 4 from biology, 3 from philosophy, 4 from mathematics/operations research, and 1 from economics (as judged by departmental affiliation). No strong correlation emerged between discipline and attitude toward the optimality heuristic, although psychologists and philosophers were most skeptical about its value. In addition, little consensus could be found about the usefulness, role, and epistemological basis of optimality principles. Nonetheless, interesting extensions (e.g., **Crow**'s), challenges, and issues were introduced. Foremost, in my view, is the need for better criteria to judge the utility and validity of optimality models. Second, the usefulness of comparing and generalizing across disciplines surfaced as an interesting issue (especially for this journal). I shall start by examining these two questions in reverse order and then I will turn to other important issues.

1. Overall concerns

Generalizability. Several commentators questioned the value or even the legitimacy of examining optimality principles across sciences. For example, **Sober** begins his commentary with, "The examples and concepts that Schoemaker cites are rather heterogeneous," and concludes with, "The way to improve our understanding of optimality concepts in science is to look at concrete examples in some detail, without assuming in advance that there is a single analysis that must work across all the cases." Others (**Cabanac, Larichev, Paelinck,** and **Toda**) prefer separate treatments of optimality across the physical, biological, and social sciences. In contrast, **Bookstein** considers all optimality principles "at best tautologies," whereas **Anderson** welcomes more extensive cross-fertilization between economics and psychology, especially in the domain of human memory (which he believes to exhibit optimal design). **Miller & Zucker** generally address the paradox that the ubiquity of optimality models runs counter to their inefficiency from a computational perspective (i.e., it can be very expensive to find and compute optimal solutions). And **Mongin** addresses the general philosophical concern that what is optimal at one level need not be optimal at a higher metalevel (resulting in an infinite regress problem).

At the level of form, methodology, and philosophy it seems useful to examine optimality principles across disciplines. Apart from potential cross-fertilization and standardization (in terminology and technique), the general perspective acknowledges that scientific methodology transcends the specifics of disciplines. In addition, it presents an opportunity to examine how the sciences interrelate and connect in the hierarchy of observations. For example, if the laws of physics do obey some deep optimality principle, then presumably animals and humans should also exhibit this fact (within a strict materialist view). The converse, however, need not hold: Perfect efficiency of economic markets does not imply perfect rationality for all players (because only a subset of traders determines the equilibrium price). Thus, from a materialist point of view, we might expect a gradient of optimality across sciences, running from the most fundamental (physics and chemistry) to higher levels (e.g., the social sciences). This gradient may in turn be countered or strengthened by additional variables or conditions that arise at higher levels (e.g., natural selection, complexity, consciousness, and perhaps free will). In terms of substance, however, it seems useful to examine optimality principles by discipline or specialty. For example, Fermat's principle of least time is quite different in its referents, process, and assumptions from, say, optimal foraging theories in ecology or capital market efficiency in economics.

Validation criteria. Several commentators address the issue of how to judge or evaluate optimality models. **Lea** especially deems it unfortunate that I "reject both predictive power and causal process" as reliable criteria. This interpretation is not what I intended. I reject each as the sole or overriding criterion for judging any theory's value. Moreover, my concern about causal accounts was raised in the context of comparing teleological and causal explanations (which **Steele** deems to be a tenuous distinction anyway). I would consider predictive power and causal understanding to be two of *several* criteria, however, for judging the validity of optimality models (as well as other kinds).

As I note later in the target article "each optimality principle . . . begs for an associated process explanation that describes causally, within the constraints of an organism or system, how it operates." It is this claim that **Miller** deems "one of Schoemaker's most central sentences," whereas **Dayan & Oberlander** cite the same sentence as an illustration of confusing levels of explanation. Both **Cabanac** and **Miller** argue explicitly for a dynamic criterion – survival or fitness for example – to judge optimality. Miller develops this view most fully, proposing two criteria: (1) dynamic adaptiveness concerning short-term behavior as the system approaches its goal(s), and (2) historical adaptation referring to long-term selective dynamics that gave rise to the system in the first place. In Miller's view, true optimality (as opposed to apparent optimality) must evolve through a trial and error process that undergoes selection and is as such exclusively applicable to dissipative systems. Although it is appealing, I find this view too narrow as it excludes the kind of "designed optimality" characteristic of the sciences of the artificial (Simon 1981).

Proposed criteria. So how *should* we judge an extremum or optimality principle in the physical (or any other) sciences? My criteria would include (1) how well it predicts or explains existing data, (2) whether it generates useful new hypotheses that are subsequently validated, (3) whether it can be confirmed by more fundamental process theories, (4) how elegant or simple it is, and (5) how computationally tractable it is. The need for a process perspective (as opposed to black box models) was especially emphasized by **Cabanac, Davison, Miller, Shafir, Stolarz-Fantino & Fantino.**

Bookstein further insists that the maximand should be independently measurable, as in the case of Fermat's principle of least time. I subsume this under criterion (1), concerning predictability (and thus implicitly its testability). In Schoemaker (1984), however, I examined Bookstein's concern in more detail using Campbell's (1970) conceptualization of a scientific theory. In Campbell's view, a scientific theory consists of two sets of propositions, one concerning the hypothesis and the other concerning the theory's dictionary. For example, consider Campbell's hypothetical theory for the law that in metals the ratio of resistance and temperature remains constant.

Let the hypothesis be that (1) u, v, and w are independent variables, (2) a and b are constants, and (3) c and d are dependent variables with $c = d$. Let the dictionary define the resistance of metals as $R = (c^2 + d^2)a$ and their temperature as $T = cd/b$. This permits us to postulate the above ratio law, since

$$R/T = (c^2 + d^2)ab/cd = 2c^2ab/cd = 2ab = \text{constant}$$

Nonetheless, this contrived theory of a correct law is without value. The reason is that the variables and constants are not independently measurable, but only in combination (as R or T). However, if u had been identified as, say, time and were also to appear in the other formulae, then this theory might be testable. For example, Boyle's gas law is valuable because its hypothesis contains propositions that are analogous to observable variables or principles. Such independent measurability is to Campbell the essential characteristic of a theory's value.

The criterion of testability can also be subsumed under the criterion of plausibility, to which the target article explicitly refers. **Byrne** examines the plausibility criterion in greater detail and finds it more a negative than positive heuristic. Because I agree with many of his observations, the above criterion set seems more appropriate. Note, however, that a full-fledged, general theory and justification of criteria to judge optimality models would require a separate article.

2. Different sciences reexamined

Physical sciences. The intriguing observation is offered by **Toda** (who in midcareer switched from physics to psychology) that physicists are optimizers in their intentions and search for truth, beauty, and simplicity but that their theories are generally not of the optimizing kind. **Sober** further questions my equation of extremum principles with those of optimality. Indeed, it is true that any assertion about optimality contains subjective, evaluative elements. What I suggested in the target article is that such principles as Fermat's may gain in stature if their maximands are plausible or desirable. It is presumably a good thing for light to travel in the least amount of time (we would all wish to do so, leaving sightseeing aside). In addition, who can argue with the beauty and utility of being efficient (which may be Nature's greatest virtue). Obviously, I am anthropomorphizing and wonder to what extent this all too human tendency affects physical scientists' judgments about the appeal and value of their extremum principles (quite apart from historical and religious remnants). It should also be noted, however, that such aesthetic appeals do not apply to all extremum principles. The law of maximum chaos (entropy) would be abhorred by most good citizens (teenagers excepted). Similarly, weather systems or balloons equalizing pressure may leave most of us quite cold.

Biology. The epistemological status of extremum principles for living systems is more difficult to assess. **Daly** suggests that even though natural selection has no goal(s) per se (I agree), it brings forth organisms that do have goals and purposes. This, in Daly's view, is the *fundamental* distinction between living and lifeless systems. Yet it remains unclear why this distinction is deep (i.e., one of kind rather than of degree). Suppose we construct an artificial world consisting of robots that undergo unnatural selection. Assume the robots exhibit genetic variance (in their programs), undergo random mutation, can mate and reproduce with heritability, and are differentially terminated depending on some arbitrary traits or behavioral features. Why, when, and how would purposes develop in these machines? And how would we recognize these purposes as being distinct from causes, reflexes, or other conditions?

Such a simulated world could nicely manipulate how much mutation is neutral – akin to Kimura's (1989) genetic drift – and to what extent selection occurs at the population, organism, and genome levels. This way we could test how well biologists who are blind with respect to the design parameters can infer the deeper laws of this ecosystem. Although Daly faults me for distinguishing various levels of selection, arguing that none of the authors cited accepts species selection, it lies at the heart of Eldredge & Gould's (1972) punctuated equilibria view of evolution. In addition, at least one of the authors cited explicitly wrote: "The entities which are subject to evolution by natural selection may not be individual organisms, but either larger entities (populations) or smaller ones (genes, or groups of genes)" (Maynard Smith 1986, p. 7). Indeed, the possibility of species or population selection seems hard to rule out. Suppose we annihilate ourselves as well as several other species in some vast nuclear war with the help of presumably "optimal" computers. Wouldn't that constitute species level selection?

The notion that through natural selection true optimizers may evolve connects with **Miller & Zucker's** general questions as to: "How it is possible that nature is computing all these optima?" They emphasize that optimal solutions are often expensive to calculate and at times impossible to find with certainty. For example, many animals – in their search for food or mates – encounter various dynamic programming problems that are known to rise exponentially in complexity as the number of variables or states increases (see Houston & McNamara 1988). Indeed, some will be NP-hard, meaning that they are not polynomially bounded, with no guaranteed closed form solutions.

One answer is provided by **Helweg & Roitblat,** who suggest that neural networks may be up to a task that has eluded operations research so far. To quote: "Although each element in the network has very limited computational capacity, a three-layer network can compute any arbitrary function" (Hecht-Nielsen 1987). Another is to consider the optimality models "as if," not to be taken literally. The interesting dimension of biological systems, however, is that true optimality can emerge (in the sense of adaptive fitness or survival). Thus, we might add to the aforementioned criteria that for biological optimality the assumptions underlying natural selection apply. That is, genetic variance, heritability, random mutation, selection pressure, and stable convergence over many generations would prima facie favor true or inherent optimality (in the sense of **Miller**'s adaptive criteria). Alternatively, we could consider this a process explanation (to accompany the optimality explanation), so that it is not so much a new criterion (relative to the earlier list) as a more specified one.

Social sciences. With respect to full consciousness, I suggest in the target article that optimality may even be more plausible because of foresight, learning, reflective intelligence, and free choice. I also note, however, that freedom of choice permits suboptimal behavior and also introduces vastly increased complexity. **Daly** argues that at the societal level, Darwinian processes are less likely to be operative. In his words,

> Sociological and anthropological functionalisms have generally failed miserably, for reasons transparent to a Darwinian: There is no rationale for expecting societies to function as if they have "purposes," because there is no process that could plausibly provide them with the sort of integrated "designs" and goal-directedness that selection gives to individual organisms.

Miller similarly questions the value of optimality arguments in economics because of the absence of cumulative selection. What is lacking, he argues, is heritable variation and increased numerosity of offspring for those who are fitter.

This criticism can be responded to at two levels: One in terms of fit with the assumptions underlying natural selection; the other by noting (as **Hyland** and **Shafir** do) that in humans optimality can also be achieved through forethought and intelligence. It is interesting that there seems to be a tradeoff between the argument for biological optimality and that for designed optimality. The less intelligent we assume humans to be, the more persuasive the biological argument will be and the less persuasive the design one. Consider, for example, the view of organizations put forth by Cyert and March (1963) or more recently by Nelson and Winter (1982). These models assume that firms have limited routines for coping with new circumstances and that new routines emerge slowly through a process of local search. In such a world, economic selection is likely to exert a considerable force. New generations of managers are burdened with the views and practices of preceding ones (a form of cultural and organizational inheritance). Consequently, they adapt poorly to new challenges, thereby permitting more suitable organizations to flourish and win out. Organizational ecology, as a subfield of sociology, explicitly adopts this view, with considerable empirical support (see Hannan & Freeman 1989). As such, bounded rationality may favor biological optimality arguments.

If, in contrast, we postulate hyperrationality, as economists often prefer, arguments centering on habit, incrementalism, and selection lose power, whereas those of designed optimality gain. Firms will hire the best minds to work on those problems that truly matter, and optimality will often be explicitly designed for. Examples include linear programming applications in oil refineries or airline scheduling, efficient portfolio allocations in financial management, or optimal forecasting or time-series. In this view, rationality is a purchaseable commodity that will flow to where it will do the most good (an instance of Coase's 1960 theorem). Firms will buy and develop intelligence to an optimal (but not maximal) degree, that is, to the point where the marginal cost of becoming still smarter equals its expected benefit. Note that in such a world surplus profit-seeking is doomed from the start.

The social sciences also suffer from ambiguity surrounding the notion of rationality. As **Wierzbicki** justly points out, rationality is not necessarily equivalent to utility maximization. First, there is the issue of cost and effort. Seemingly suboptimal decision rules, for example, Simon's (1957) satisficing or other noncompensatory screening methods may appear optimal when one takes cognitive and computational costs into account (see Johnson & Payne 1985). Since some of these costs are unobservable or hard to measure, they can become powerful fudge factors in ex post facto rationalizations. Second, philosophical disagreements remain over different kinds of rationality and the appropriate criteria to adjudicate them. Some theorists require no more than subjective rationality (just within the person's worldview and beliefs), whereas others favor adaptive rationality (linked to survival) or procedural rationality (adapted to our cognitive limits).

3. Human optimality

Economic man. The complexities of wielding the sword of optimality wisely (to borrow **Wierzbicki**'s metaphor) are especially noticeable in the rational-economic man tradition. For nearly a century, economists have faithfully followed Pareto's (1897) advice that "once we have obtained a photograph of his taste, . . . the individual may disappear." Being among the most advanced of the social sciences, economics has reduced man to a mere subscript in mathematical equations (see Georgescu-Roegen 1971). Having no choice but to maximize his expected utility (Scitovsky 1976), basic preferences are the only thing researchers need to predict economic man's behavior. [See also Caporael et al.: "Selfishness Examined: Cooperation in the Absence of Egoistic Incentive" *BBS* 12(4) 1989.]

This robotic image, further endowed with nearly infinite problem solving capacity, has apparently served the masters of rationality well. Their brain child (i.e., the fiction called economic man) has become almost as rational as the brightest of his economic parents. For example, whereas initially he could only "maximize" his utility under fixed search rules (Stigler 1961), today he knows how to use "more optimal" variable search rules (Rothschild 1974). Or, whereas once he could be fooled by governmental policies entailing monetary and fiscal levers, today his rational expectations largely nullify these regulatory interventions (Lucas 1980).

The above caricature of economic man is not meant to belittle pathbreaking work in economics, but to highlight the inherent difficulties encountered when using optimality principles in the social realm. The first dilemma concerns the discrepancy between the IQ of economic man now versus his IQ even a few decades ago. The second dilemma is that as economic man becomes more firmly rooted, and indeed makes colonizing excursions into neighboring terrain (law, crime, animal behavior, etc.), his sister discipline of psychology is mounting a vigorous attack on the behavioral validity of the underlying assumptions (see Hogarth & Reder 1986).

The standard reply revolves around issues of comparative advantage, methodological preference (positivism), and the lack of strong alternative theories. Psychological man, it is usually countered, is still too frail, too unaxiomatized, and just not "together enough" to shoulder the burden of mathematical aggregation from individual to market behavior. It is presumed that such aggregation should follow the same path of mathematical rigor, generality, and tractability that characterizes the elegant closed-form development of economic man. The real problem, however, is that economic man is limping. Somehow his sterile development, in an environment devoid of real-world contamination, has resulted in an ill-adjusted adult, whose beauty rests mostly in the eyes of his parents.

Decision sciences. If we abandon the "as if" perspective and get down to actual, deliberate optimization of real-world problems, additional challenges arise. As **Miller & Zucker** emphasize, many of life's optimization problems seem hopelessly complex, although Clark (target article, this issue) offers some rays of hope. Furthermore, insofar as human preferences or beliefs are to be measured, optimization sciences run into precisely the problems they are meant to cure, namely, bounded rationality. This problem is especially acute in decision analysis, which requires utility functions as one of its subjective inputs. The problem is that of obtaining preference judgments that satisfy the axioms (e.g., those of von Neumann & Morgenstern 1947) on which the entire normative apparatus rests.

To illustrate, let me mention briefly some of my own work with John C. Hershey on utility measurement. We ask subjects to provide a certainty equivalence (CE) judgment for a simple gamble – a 50-50 chance of getting $0 or $200, for example. Suppose our subject provides a CE of $80 (i.e., being indifferent to getting $80 or the gamble). A week or so later, we present the subject with a choice between $80 for sure versus a 40% chance of getting $200 and a 60% of getting $0. This time, however, we ask for a probability equivalence (PE), that is, an adjustment of the 40% chance to a level that brings about indifference. Expected utility (EU) theory, as well as any holistic choice model with a one-to-one mapping between money and utility, would predict an answer of 50% (because that was the initial indifference state). In our experiments, we find that significantly more subjects give a response above than below 50% for this and similar problems (see Hershey & Schoemaker 1985). In subsequent work (Schoemaker & Hershey 1991), we trace such CE-PE discrepancies to PE reframing, regression toward the mean, and anchoring with insufficient adjustment (see also Johnson & Schkade 1989).

Probability equivalence (PE) reframing here refers to subjects translating the gamble downward in the PE mode, as though it were a choice between $0 for sure versus a 40% chance of gaining $100 and a 60% chance of losing $100. That is, outcomes are recoded relative to the sure amount, which serves as the new reference point or status quo. This happens only in the PE mode where the sure amount is kept fixed (whereas it varies in the CE mode). The psychological shift from the gain to the mixed payoff domain typically induces an increase in risk-aversion, compatible with prospect theory's value function shape (see Kahneman & Tversky 1979). Without understanding such psychological evaluation biases, it becomes very difficult to arrive at reliable utility measurements. Unfortunately, CE-PE discrepancies are just one of several potential obstacles in utility encoding (see Hershey et al. 1982).

The point to be made is that designing for optimality is by no means simple, for mathematical/computational, philosophical, and psychological reasons (see Kleindorfer et al. 1991 for elaborations). So whereas I would agree with **Baron** and **Rinnooy Kan** about the potential value of optimality principles in providing normative guidance, I and others (e.g., **Davison** and **Levine**) are less sanguine about their current value (see also Bell et al. 1989). Questions of uniqueness (or decisiveness), lack of consensus and limited operationality continue to plague the engineering of rational choice (see also March 1988).

In summary, the optimality game is a difficult one to play well, both descriptively and prescriptively. Unlike some other research programs, it can easily degenerate into tautology or can otherwise retard scientific progress. In the instrumentalist view of **Sahlin,** optimality may be little more than a tool of science, but as **Mazur** and others have emphasized, it can easily be abused.

References

Ainslie, G. W. (1974) Impulse control in pigeons. *Journal of the Experimental Analysis of Behavior* 21:485–89. [JEM]
(1985) Beyond microeconomics: Conflict among interests in a multiple self as a determinant of value. In: *The multiple self*, ed. J. Elster. Cambridge University Press. [JEM]
Alchian, A. A. (1950) Uncertainty, evolution and economic theory. *Journal of Political Economy* 58:211–21. [aPJHS]
Ancot, J.-P. & Paelinck, J. H. P. (1983) The spatial econometrics of the European FLEUR-model. In: *Evolving geographical structures*, ed. D. Griffith & A. Lea. Martinus Nijhoff Publishers. [JHPP]
Anderson, J. R. (1990a) *The adaptive character of thought*. Erlbaum. [JRA]
(1990b) *Cognitive psychology and its implications*. 3rd ed. W. H. Freeman. [JRA]
Anderson, J. R. & Milson, R. (1989) Human memory: An adaptive perspective. *Psychological Review* 96(4):703–19. [JRA]
Arrow, K. J. (1951) *Social choice and individual values*. John Wiley. [aPJHS]
Ashby, W. R. (1956) *Introduction to cybernetics*. John Wiley. [aPJHS]
Axelrod, R. (1984) *The evolution of cooperation*. Basic Books. [APW]
Baddeley, A. (1986) *Working memory*. Oxford University Press. [JRA]
Barigozzi, C. (1980) *Vito Volterra symposium on mathematical models in biology*. Springer-Verlag. [aPJHS]
Barney, J. B. & Ouchi, W. G. (1986) *Organizational economics*. Jossey-Bass. [aPJHS]
Baron, J. (1988) *Thinking and deciding*. Cambridge University Press. [JB]
Baron, J. & Brown, R. V. (in press) *Teaching decision making to adolescents*. Erlbaum. [JB]
Barrow, J. D. & Tipler, F. J. (1986) *The anthropic cosmological principle*. Cambridge University Press. [GFM]
Bateson, G. (1979) *Mind and nature*. Wildwood House, Ltd. [aPJHS]
Battalio, R. C., Kagel, J. H., Rachlin, H. & Green, L. (1981) Commodity-choice behavior with pigeons as subjects. *Journal of Political Economy* 89:1. [aPJHS]
Beatty, J. (1987) *Natural selection and the null hypothesis*. In: *The latest on the best*, ed. J. Dupré. MIT Press. [aPJHS]
Becker, G. S. (1976) *The economic approach to human behavior*. University of Chicago Press. [aPJHS, JEM]
Becker, G. S. & Murphy, K. M. (1988) A theory of rational addiction. *Journal of Political Economy* 96:675–99. [JEM]
Bell, D., Raiffa, H. & Tversky, H. (1988) *Decision making*. Cambridge University Press. [rPJHS]
Blake, A. & Zisserman, A. (1988) *Visual reconstruction*. MIT Press. [DAM]
Boland, L. A. (1981) On the futility of criticizing the neoclassical maximization hypothesis. *American Economic Review* 71(5):1031–36. [aPJHS]
Bordley, R. F. (1983) A central principle of science: Optimization. *Behavioral Science* 28:53–64. [aPJHS]
Boulding, K. E. (1962) *Conflicts and defense*. Harper & Row. [aPJHS]
Bounds, D. G. (1987) New optimization methods from physics and biology. *Nature* 329:215–20. [aPJHS]
Bradley, J. (1971) *Mach's philosophy of science*. University of London Press. [aPJHS]
Braithwait, R. B. (1966) Causal and teleological explanation. In: *Purpose in nature*, ed. J. V. Canfield. Prentice Hall. [aPJHS]
Brescia, F., Arents, F., Meislich, H. & Turk, A. (1966) *Fundamentals of chemistry*. Academic Press. [aPJHS]
Bridgman, P. W. (1948) *The logic of modern physics*. MacMillan. [aPJHS]
Bromberger, S. (1970) Why-questions. In: *Readings in the philosophy of science*, ed. B. Brody, Prentice-Hall. [aPJHS]
Buchanan, J. M. & Tullock, G. (1962) *The calculus of consent*. University of Michigan Press. [aPJHS]
Bull, J. J. (ed.) (1983) *Evolution of sex-determining mechanisms*. Benjamin/Cummings. [aPJHS]
Bullock, D. & Grossberg, S. (1988) Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review* 95:49–90. [DSL]
Bunge, M. (1979) *Causality and modern science*. Harvard University Press. [aPJHS]
Burrell, Q. L. & Cane, V. R. (1982) The analysis of library data. *Journal of the Royal Statistical Society, Series A*(145):439–71. [JRA]
Cabanac, M. & LeBlanc, J. (1983) Physiological conflict in humans: Fatigue *vs.* cold discomfort. *American Journal of Physiology* 244:R621–28. [MC]
Callahan, J. J. (1976) The curvature of space in a finite universe. *Scientific American*, August:90–99. [aPJHS]
Camerer, C. F. (1990) Recent tests of generalizations of expected utility theory (manuscript). Department of Decision Sciences, The Wharton School, University of Pennsylvania, Philadelphia, PA. [ESh]
Campbell, N. (1970) What is a theory? In: *Readings in the philosophy of science*, ed. B. A. Brody. Prentice-Hall. [rPJHS]
Canfield, J. V. (1966) *Purpose in nature*. Prentice-Hall. [aPJHS]
Carver, C. S. & Scheier, M. F. (1982) Control theory: A useful conceptual framework for personality-social, clinical, and health psychology. *Psychological Bulletin* 92:111–35. [MEH]
Chandrasekhar, S. (1987) *Truth and beauty: Aesthetics and motivations in science*. University of Chicago Press. [APW]
Charnov, E. L. (1976) Optimal foraging: The marginal value theorem. *Theoretical Population Biology* 9:126–43. [aPJHS]
Chicago Tribune (1988) Article on 1988 Nobel Prize winners in physics, Oct. 20, p. 2. [aPJHS]
Chisholm, R. M. (1956) Sentences about believing. *Proceedings of the Aristotelian Society* 56:125–40. [aPJHS]
Chomsky, N. (1980) *Rules and representation*. Columbia Press. [aPJHS]
Clark, C. W. (1991) Modeling behavioral adaptations. *Behavioral and Brain Sciences* 14(1):85–117. [rPJHS, DAH]
Coase, R. H. (1960) The problem of social cost. *Journal of Law and Economics* 3:1–45. [arPJHS]
Cohen, L. J. (1981) Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences* 4:317–70. [aPJHS]
Coleman, J. S. (1990) *Foundations of social theory*. Harvard University Press. [APW]
Coombs, C. H., Dawes, R. M. & Tversky, A. (1970) *Mathematical psychology: An introduction*. Prentice-Hall. [aPJHS]
Cooper, W. S. (1987) Decision theory as a branch of evolutionary theory: A biological derivation of the savage axioms. *Psychological Review* 94(4):395–411. [aPJHS]
Cosmides, L. M. & Tooby, J. (1981) Cytoplasmic inheritance and intragenomic conflict. *Journal of Theoretical Biology* 89:83–129. [MDal]
(1987) From evolution to behavior: Evolutionary psychology as the missing link. In: *The latest on the best: Essays on evolution and optimality*, ed. J. Dupre. MIT Press. [GFM]
Cournot, A. (1838/1971) *Researches into the mathematical principles of the theory of wealth*. Augustus M. Kelly. [aPJHS]
Cyert, R. & March, J. (1963) *A behavioral theory of the firm*. Prentice-Hall. [rPJHS]
D'Arcy Thompson, W. (1917/1942) *On growth and form*. Cambridge University Press. [aPJHS]
Darwin, C. (1859) *On the origin of species*. J. Murray. [aPJHS]
Davidson, D. (1986) *Essays on actions and events*. Oxford University Press. [aPJHS]
Davison, M. & Kerr, A. (1989) Sensitivity of time allocation to an overall reinforcer-rate feedback function in concurrent interval schedules. *Journal of the Experimental Analysis of Behavior* 51:215–31. [MD]
Dawkins, R. (1982) *The extended phenotype*. W. H. Freeman. [MDal]
(1988) *The blind watchmaker*. Penguin Books. [JHPP]
de Chardin, P. T. (1968) *Future of man*. Fontana. [MEH]
De Finetti, B. (1975) *Theory of probability*. John Wiley & Sons. [aPJHS]
Dennett, D. C. (1978) *Brainstorms: Philosophical essays on mind and psychology*. Bradford Books. [aPJHS]
(1987) *The intentional stance*. MIT Press/Bradford Books. [PD]
d'Espagnat, B. (1979) The quantum theory and reality. *Scientific American*, vol. 241. [aPJHS]
Downs, A. (1957) *An economic theory of democracy*. Harper. [aPJHS]
Dupré, J. (ed.) (1987) *The latest on the best: Essays on evolution and optimality*. MIT Press. [aPJHS]
Eatwell, J., Milgate, M. & Newman, P. (1988). *The new Palgrave: A dictionary of economics*. Macmillan. [aPJHS]
Edwards, P. (ed.) (1972) *The encyclopedia of philosophy*, vol. 1–8. Macmillan. [aPJHS]
Einhorn, H. & Hogarth, R. M. (1981) Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology* 32:53–88. [aPJHS]
(1986) Judging probable cause. *Psychological Bulletin* 99(1):3–19. [aPJHS]
Eldredge, N. & Gould, J. (1972) Punctuated evolution: An alternative to phyletic gradualism. In: *Models of paleobiology*, ed. T. Schopf. Freeman. [rPJHS]
Ellsberg, D. (1961) Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics* 75:643–99. [JB]
Elster, J. (1979) *Ulysses and the sirens*. Cambridge University Press (Cambridge). [aPJHS]
(1982) *Explaining technical change*. Cambridge University Press (Cambridge). [aPJHS]
(1983) Sour grapes. In: *Studies in the subversion of rationality*. Cambridge University Press (Cambridge). [aPJHS]
(1989) *Solomonic judgments: Studies in the limitations of rationality*. Cambridge University Press (Cambridge). [aPJHS]

Euler, L. (1744/1952). Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, Lausanne. Reprinted in *Leonhardi Euleri Opera Omnia*. vol. 24, Series 1 Bern. [aPJHS]
Evans, J. St. B. T. (1989) *Bias in human reasoning*. Erlbaum. [MEH]
Fama, E. F. (1976) *Foundations of finance*. Basic Books. [aPJHS]
Fama, E. F. & Miller, M. H. (1972) *The theory of finance*. Holt, Rinehart & Winston. [aPJHS]
Fantino, E. & Abarca, N. (1985) Choice, optimal foraging, and the delay-reduction hypothesis. *Behavioral and Brain Sciences* 8:315–62. [aPJHS, SS-F]
Fantino, E. & Preston, R. A. (1988a) Foraging for integration. *Behavioral and Brain Sciences* 11:683–84. [SS-F]
(1988b) Choice and foraging: The effects of accessibility on acceptability. *Journal of the Experimental Analysis of Behavior* 50:395–403. [SS-F]
Farris, S. (1983) The logical basis of phylogenetic analysis. In: *Advances in cladistics*, 2nd ed., ed. N. Platnick & V. Funk. Columbia University Press. Also in E. Sober, ed. (1984) *Conceptual issues in evolutionary biology*. MIT Press. [ESo]
Feynman, R. P., Leighton, R. B. & Sands, M. (1964) *The Feynman lectures of physics*, vol. 1. Addison-Wesley. [aPJHS, FLB]
Fisher, R. A. (1930/1958) *The genetical theory of natural selection*. The Clarendon Press. Revised edition. Dover Publications. [JFC]
Flash, T. & Hogan, N. (1985) The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience* 5:1688–1703. [DSL]
Fodor, J. A. (1975) *The language of thought*. Harvester Press. [PD]
(1981) Three cheers for propositional attitudes! In: *Representations*. Harvester Press. [PD]
Foster, C. (1990) Algorithms, abstraction and implementation. PhD. thesis, Centre for Cognitive Science, University of Edinburgh. [PD]
Friedman, M. (1953) *Essays in positive economics*. University of Chicago Press. [aPJHS]
Frisch, D. & Baron, J. (1988) Ambiguity and rationality. *Journal of Behavioral Decision Making* 1:149–57. [JB]
Garey, M. R. & Johnson, D. S. (1979) *Computers and intractability*. W. H. Freeman & Co. [DAM]
Georgescu-Roegen, N. (1971) *The entropy law and the economic process*. Harvard University Press. [rPJHS]
Ghiselin, M. T. (1974) *The economy of nature and the evolution of sex*. University of California Press. [aPJHS]
Göttinger, H. W. (1982) Computational cost and bounded rationality. In: *Philosophy of economics*, ed. W. Stegmüler, W. Balzer & W. Spohn. Springer. [PM]
Gould, S. J. (1980) *The panda's thumb*. Norton. [DSL]
Gould, S. J. & Lewontin, R. C. (1979) The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptionist programme. *Proceedings of the Royal Society B* 205:591–98. [aPJHS, DAH, KMS, MDal, PD]
Gray, J. A. & Smith, P. T. (1969) An arousal-decision model for partial reinforcement and discrimination learning. In: *Animal discrimination learning*, ed. R. M. Gilbert & N. S. Sutherland. Academic Press. [DSL]
Grether, D. M. & Plott, C. R. (1979) Economic theory of choice and the preference reversal phenomenon. *American Economic Review* 69(4):623–38. [aPJHS]
Griffin, D. R. (1981) *The question of animal awareness*. Rockefeller University Press. [aPJHS]
Grossberg, S. & Gutowski, W. (1987) Neural dynamics of decision making under risk: Affective balance and cognitive-emotional interactions. *Psychological Review* 94:300–18. [DSL]
Grossberg, S. & Mingolla, E. (1985) Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review* 92:173–211. [DSL]
Hacking, I. (1983) *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge University Press. [N-ES]
Haldane, J. B. S. (1927) *Possible worlds*. Chatto & Windus. [aPJHS]
Hamilton, W. R. (1834/1940) On a general method in dynamics. *Philosophical Transactions of the Royal Society*. Also in *The mathematical papers of Sir William Rowan Hamilton*, vol. 2, Dynamics. Cambridge University Press. [aPJHS]
(1835/1940) Second essay on a general method in dynamics. *Philosophical Transactions of the Royal Society*. (Also in *The mathematical papers of Sir William Rowan Hamilton*, vol. 2, Dynamics. Cambridge University Press.) [aPJHS]
Hannan, M. & Freeman, J. (1989) *Organizational ecology*. Harvard University Press. [rPJHS]
Hansson, B. (1988) Risk aversion as a problem of conjoint measurement. In: *Decision, probability, and utility*, ed. P. Gärdenfors & N.-E. Sahlin. Cambridge University Press. [N-ES]
Hecht-Nielsen, R. (1987) Kolmogorov's mapping neural network existence theorem. *Proceedings of the IEEE First International Conference on Neural Networks*, San Diego, CA, vol. 3. [rPJHS, DAH]
Heiner, R. A. (1983) The origin of predictable behavior. *American Economic Review* 73:560–85. [DSL]
Heisenberg, W. (1955) *Das Naturbild der Heutige Physik*. Hamburg. [aPJHS]
Herrnstein, R. J. (1990) Rational choice theory: Necessary but not sufficient. *American Psychologist* 45:356–67. [MD]
Herrnstein, R. J. & Mazur, J. E. (1987) Making up our minds: A new model of economic behavior. *The Sciences* 27:40–47. [JEM]
Herrnstein, R. J. & Vaughan, W. (1980) Melioration and behavioral allocation. In: *Limits to action: The allocation of individual behavior*, ed. J. E. R. Staddon. Academic Press. [MD]
Hershey, J. C. & Schoemaker, P. J. H. (1985) Probability vs. certainty equivalence methods. In: Utility measurement: Are they equivalent? *Management Science* 31:1213–31. [rPJHS]
Hershey, J. C., Kunreuther, H. C. & Schoemaker, P. J. H. (1982) Sources of bias in assessment procedures for utility functions. *Management Science* 28:936–54. [arPJHS]
Hess, E. H., (1973) *Imprinting*. D. Van Nostrand. [aPJHS]
Hesse, M. (1969) Simplicity. In: *The encyclopedia of philosophy*, ed. P. Edwards. Macmillan. [ESo]
Hirshleifer, J. (1977) Economics from a biological point of view. *Journal of Law and Economics* 20:1–52. [aPJHS, SEGL]
(1987) *Economic behavior in adversity*. University of Chicago Press. [aPJHS]
Hobbes, T. (1909) *Leviathan*. The Clarendon Press. [aPJHS]
Hogarth, R. M. & Reder, M. W., eds. (1986) The behavioral foundations of economic theory. *Journal of Business* 59(4) Part 2. [arPJHS, ESh]
Houston, A. I. & McNamara, J. M. (1987) Singing to attract a mate: A stochastic dynamic game. *Journal of Theoretical Biology* 129:57–68. [DAH]
(1988) A framework for the functional analysis of behaviour. *Behavioral and Brain Sciences* 11:117–63. [rPJHS, DAH]
Houthakker, H. S. (1956) Economics and biology: Specialization and speciation. *Kyklos* 9:181–89. [aPJHS]
Hubel, D. H. & Wiesel, T. (1979) Brain mechanisms of vision. *Scientific American*, September. [aPJHS]
Hull, C. L. (1930) Knowledge and purpose as habit mechanisms. *Psychological Review* 37:511–25. [KMS]
(1931) Goal attraction and directing ideas conceived as habit phenomena. *Psychological Review* 38:487–506. [KMS]
(1943) *Principles of behavior*. Appleton Century. [KMS]
(1952) *A behavior system: An introduction to behavior theory concerning the individual organism*. Yale University Press. [KMS]
Hume, D. A. (1988) *A treatise on human nature*, ed. L. A. Selby-Bigge.
Hume, A. L. & Irwin, R. J. (1974) Bias functions and operating characteristics of rats discriminating auditory stimuli. *Journal of the Experimental Analysis of Behavior* 21:285–95. [MD]
Hyland, M. E. (1988) Motivational control theory: An integrative framework. *Journal of Personality and Social Psychology* 55:642–51. [MEH]
Hylleraas, E. A. (1970) *Mathematical and theoretical physics*, vol. 1. John Wiley & Sons. [aPJHS]
Jantsch, E. & Waddington, C. H. (1976) *Evolution and consciousness*. Addison-Wesley. [aPJHS]
Johnson, E. & Payne, J. (1985) Effort and accuracy in choice. *Management Science* 30:395–414. [rPJHS]
Johnson, E. J. & Schkade, D. A. (1989) Bias in utility assessments: Further evidence and explanations. *Management Science* 35:406–24. [rPJHS]
Jungermann, H. (1983) The two rationality camps. In: *Decision-making under uncertainty*, ed. R. W. Scholz. North-Holland. [OL]
Kahneman, D. & Tversky, A. (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–91. [arPJHS, JB]
(1984) Choices, values, and frames. *American Psychologist* 39:341–50. [JB]
Kahneman, D., Slovic, P. & Tversky, A. (1982) *Judgement under uncertainty: Heuristics and biases*. Cambridge University Press. [aPJHS]
Kamil, A. (1978). Systematic foraging by a nectar feeding bird: The amakihi (*Loxops virens*). *Journal of Comparative and Physiological Psychology* 92:388–96. [DAH]
Kamil, A. C. & Roitblat, H. L. (1985) Foraging theory: Implications for animal learning and cognition. *Annual Review of Psychology* 36:141–69. [DAH]
Karlin, S. & Lessard, S. (1986) *Theoretical studies on sex ratio evolution*. Princeton University Press. [aPJHS]

Keeney, R. L. & Raiffa, H. (1976) *Decisions with multiple objectives: Preferences and value tradeoffs.* John Wiley & Sons. [aPJHS]
Kim, J. (1981) Causes as explanations: A critique. *Theory and Decision* 13:293–309. [aPJHS]
Kimura, M. (1979) The neutral theory of molecular evolution. *Scientific American* vol. 241. [aPJHS]
(1983) *The neutral theory of molecular evolution.* Cambridge University Press. [aPJHS]
(1989) The neutral theory of molecular evolution and the world view of neutralists. *Genome* 31:24–31. [rPJHS]
Kingsland, S. E. (1985) *Modeling nature.* University of Chicago Press. [aPJHS]
Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983) Optimization by simulated annealing. *Science* 220:671–79. [aPJHS, DAH, DSL]
Kirousis, L. M. & Papadimitriou, C. H. (1988) The complexity of recognizing polyhedral scenes. *Journal of Computer and System Sciences* 37:14–38. [DAM]
Kitcher, P. (1985) *Vaulting ambition: Sociobiology and the quest for human nature.* MIT Press. [aPJHS]
(1987) Why not the best? In: *The latest on the best,* ed. J. Dupré. MIT Press. [aPJHS]
Kleidon, A. W. (1986) Anomalies in financial economics: Blue print for change. *Journal of Business* 59, 4(2):5451–68. [aPJHS]
Kleindorfer, P., Kunreuther, H. & Schoemaker, P. (1991) *Decision sciences: An integrative perspective.* Cambridge University Press (in press). [rPJHS]
Klopf, A. H. (1982) *The hedonistic neuron.* Hemisphere. [DSL]
(1988) A neuronal model of classical conditioning. *Psychobiology* 16:85–125. [DSL]
Koopmans, T. C. (1972) Representation of preference orderings over time. In: *Decisions and organizations,* ed. C. B. McGuire & T. R. Radner. North-Holland. [aPJHS]
Krebs, J. R. & Davies, N. B., eds. (1978) *Behavioral ecology: An evolutionary approach.* Sinauer. [JEM]
Krebs, J. R., Kacelnik, A. & Taylor, P. J. (1978) Test of optimal sampling by foraging great tits. *Nature* 275:27–31. [DAH]
Kuhn, T. S. (1962) *The structure of scientific revolutions.* University of Chicago Press. [aPJHS]
Lagrange, L. (1788/1853) *Mécanique analytique, Paris.* (Also 3rd ed., ed. M. J. Bertrand. Mallet-Bachelier.) [aPJHS]
Lakatos, I. (1970) Falsification and the methodology of scientific research programmes. In: *Criticism and the growth of knowledge,* ed. I. Lakatos & A. Musgrave. Cambridge University Press (Cambridge). [aPJHS]
Lanczos, C. (1970) *The variational principles of mechanics,* 4th ed. University of Toronto Press. [FLB]
Langton, C. (1990) Computation at the edge of chaos: Phase transitions and emergent computation. T-13 and Center for Nonlinear Studies, Los Alamos National Laboratory, Technical Report, LA-UR-90-379. [GFM]
Larichev, O. & Moshkovich, H. (1988) Limits to decision-making ability in direct multiattribute alternative evaluation. *Organizational Behavior and Human Decision* 42:217–33. [OL]
Larrick, R. P., Morgan, J. N. & Nisbett, R. E. (in press). Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science.* [JB]
Layzer, D. (1975) The arrow of time. *Scientific American* December:56–69. [aPJHS]
Lea, S. E. G. (1984) *Instinct, environment and behaviour.* Methuen. [SEGL]
Lea, S. E. G., Tarpy, R. M. & Webley, P. (1987) *The individual in the economy.* Cambridge University Press. [SEGL]
Lenoir, T. (1982) *The strategy of life: Teleology and mechanics in nineteenth century German biology.* D. Reidel. [aPJHS]
Leven, S. J. & Levine, D. S. (1987) Effects of reinforcement on knowledge retrieval and evaluation. *First International Conference on Neural Networks,* San Diego, CA. IEEE/ICNN, vol. 2. [DSL]
Levine, D. S. (1983) Neural population modeling and psychology: A review. *Mathematical Biosciences* 66:1–86. [DSL]
Levine, D. S. & Grossberg, S. (1976) Visual illusions in neural networks: Line neutralization, tilt after-effect, and angle expansion. *Journal of Theoretical Biology* 61:477–504. [DSL]
Levine, D. S. & Leven, S. J., eds. (1990) *Motivation, emotion, and goal direction in neural networks.* Erlbaum. [DSL]
Levine, D. S. & Prueitt, P. S. (1989) Modeling some effects of frontal lobe damage: Novelty and perseveration. *Neural Networks* 2:103–16. [DSL]
Lewin, K. (1943) Defining the "field at a given time." *Psychological Review* 50:292–310. [MEH]
Lewontin, R. C. (1974) *The genetic basis of evolutionary change.* Columbia University Press. [aPJHS]
Loewenstein, G. (1987) Anticipation and the valuation of delayed consumption. *The Economic Journal* 97:666–84. [aPJHS]
Logan, F. A. (1965) Decision making by rats: Delay versus amount of reward. *Journal of Comparative and Physiological Psychology* 59:1–12. [JEM]
Loomes, G. & Sugden, S. (1982) Regret theory: An alternative approach to rational choice under uncertainty. *Economic Journal* 92:805–24. [aPJHS]
Losee, J. (1980) *A historical introduction to the philosophy of science,* 2nd ed. Oxford University Press. [N-ES]
Lotka, A. J. (1956) *Elements of mathematical biology.* Dover. [aPJHS]
Lucas, R. E. (1980) Two illustrations of the quantity theory of money. *American Economic Review* 70:1005–14. [rPJHS]
Luce, R. D. & Raiffa, H. (1957) *Games and decisions.* Wiley. [aPJHS]
Mach, E. (1883/1942) *The science of mechanics.* Open Court. [aPJHS]
Machina, M. (1982) Expected utility analysis without the independence axiom. *Econometrica* 50:277–323. [aPJHS, ESh]
Mackie, J. L. (1980) *The cement of the universe: A study of causation.* Oxford University Press (Oxford). [aPJHS]
Magill, M. J. P. (1970) *On a general economic theory of motion.* Springer-Verlag. [aPJHS]
Malkiel, B. G. (1989) Is the stock market efficient? *Science* 243:1313–18. [aPJHS]
March, J. G. (1978) Bounded rationality, ambiguity and the engineering choice. *Bell Journal of Economics* 9:587–608. [aPJHS]
(1988) *Decision and organizations.* Basil Blackwell. [rPJHS]
Marr, D. (1982) *Vision.* Freeman. [JRA, PD]
Maupertuis, P. L. M. (1744) Accord de differentes loix de la nature, qui avoient jusqu'ici paru incompatibles. *Mémoires de l'Académie des Sciences de Paris,* April 15. [aPJHS]
Maynard Smith, J. (1974) The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology* 47:209–21. [PD]
(1978) Optimization theory in evolution. *Annual Review of Ecological Systems* 9:31–56. [aPJHS, DAH]
(1982) *Evolution and the theory of games.* Cambridge University Press (Cambridge). [aPJHS]
(1984) Game theory and the evolution of behaviour. *Behavioral and Brain Sciences* 7(1):95–100. [aPJHS]
(1986) *The problems of biology.* Oxford University Press. [rPJHS]
Mayr, E. (1963) *Animal species and evolution.* Harvard University Press. [aPJHS]
(1983) How to carry out the adaptationist program. *American Naturalist* 121:324–34. [aPJHS]
Mazur, J. E. (1981) Optimization theory fails to predict performance of pigeons in a two-response situation. *Science* 214:823–25. [aPJHS]
(1987) An adjusting procedure for studying delayed reinforcement. In: *Quantitative analyses of behavior, vol. 5, The effect of delay and of intervening events on reinforcement value,* ed. M. L. Commons, J. E. Mazur, J. A. Nevin & H. Rachlin. Erlbaum. [JEM]
(1988) Choice between small certain and large uncertain reinforcers. *Animal Learning & Behavior* 16:199–205. [JEM]
McCarthy, D. & Davison, M. (1984) Isobias and alloiobias functions in animal psychophysics. *Journal of Experimental Psychology: Animal Behavior Processes* 10:390–409. [MD]
McClelland, J. L., Rumelhart, D. E. & the PDP research group (1986) *Parallel distributed processing: Explorations in the microstructure of cognition.* I. Bradford Books. [JRA]
McClosky, D. (1985) *The rhetoric of economics.* University of Wisconsin Press. [FLB]
McFarland, D. J. (1977) Decision making in animals. *Nature* 269, Sept. 1. [aPJHS]
Miller, D. A. & Zucker, S. W. (1990) Copositive-plus Lemke algorithm solves polymatrix games. Technical paper, McGill Research Centre for Intelligent Machines. [DAM]
Miller, M. (1986) Behavioral rationality in finance: The case of dividends. *Journal of Business* 59, 4(2):5451–68. [aPJHS]
Milnor, J. (1954) Games against nature. In: *Decision Processes,* ed. R. M. Thrall, C. H. Coombs & R. L. Davis. Wiley. [aPJHS]
Mongin, P. (1986) Are "all-and-some" statements falsifiable after all? The example of utility theory. *Economics and Philosophy* 2:185–95. [aPJHS]
Nagel, E. (1953) Teleological explanations and teleological systems. In: *Vision and action,* ed. S. Ratner. Rutgers University Press. [aPJHS]
(1961) *The structure of science.* Harcourt, Brace & World. [aPJHS]
Nagylaki, T. (1989) Rate of evolution of a character without epistasis. *Proceedings of the National Academy of Sciences, USA* 86:1910–13. [JFC]

Nelson, R. & Winter, S. (1982) *An evolutionary theory of economic change.* Harvard University Press. [rPJHS, APW]

Newell, A. & Simon, H. W. (1972) *Human problem solving.* Prentice-Hall. [aPJHS]

Nisbett, R. & Ross, L. (1980) *Human inference: Strategies and shortcomings of social judgment.* Prentice-Hall. [SS-F]

Oppenheimer, R. (1956) Analogy in science. *American Psychology* 11. [aPJHS]

Oster, G. F. & Wilson, E. O. (1978) A critique of optimization theory in evolutionary biology. In: *Caste and ecology in the social insects.* Princeton University Press. [aPJHS]

Paelinck, J. H. P. (1990) Vreemde dynamiek in de scheikunde en de ruimtelijke economie (Strange dynamics in chemistry and spatial economics). In: *Schetsen van een tijdperk* (Dedicatus Prof. dr B. Leijnse), ed. B. G. Blijenberg & G. J. M. Boerma. Erasmus University Press. [JHPP]

Paelinck, J. H. P. & Vossen, P. H. (1984) *The quest for optimality.* Gower. [JHPP]

Paltridge, G. W. (1975) Global dynamics and climate – A system of minimum entropy exchange. *Quarterly Journal of the Royal Meteorological Society* 101:475–84. [aPJHS]

Pareto, V. (1897) The new theories of economics. *Journal of Political Economy* 5:485–502. [arPJHS]

Parker, G. A. (1978) Searching for mates. In: *Behavioral ecology: An evolutionary approach,* ed. J. R. Krebs & N. B. Davies. Blackwell. [MC]

Partington, J. R. (1964) *A history of chemistry.* Macmillan & Co., Ltd. [aPJHS]

Pen, J. (1988) Boekbesprekingen: Review of S. E. G. Lea, R. M. Tarpy & P. Webley "The Individual in the Economy." *De Economist* 136:403–05. [SEGL]

Pepper, C. S. (1942) *World hypotheses.* University of California Press. [aPJHS]

Piaget, J. & Inhelder, B. (1975) *The origin of the idea of chance in children.* Norton. [aPJHS]

Pierce, G. J. & Ollason, J. G. (1987) Eight reasons why optimal foraging theory is a complete waste of time. *Oikos* 49:111–17. [DAH]

Popper, K. (1959) *The logic of scientific discovery.* Hutchinson. [ESo]

(1968) *Conjectures and refutations: The growth of scientific knowledge.* Harper & Row. [aPJHS]

Posner, R. (1977) *Economic analysis of law.* Little, Brown. [aPJHS]

Powers, W. T. (1978) Quantitative analysis of purposive systems. *Psychological Review* 85:417–35. [MEH]

Press, F. & Siever, R. (1974) *Earth.* W. H. Freeman & Co. [aPJHS]

Prigogine, I. (1980) *Physique, temps et devenir.* Masson. [JHPP]

Prigogine, I. & Stengers, I. (1984) *Order out of chaos.* Bantam Books. [aPJHS, GFM]

Quine, W. (1966) Simple theories of a complex world. In: *The ways of paradox and other essays.* Random House. [ESo]

(1969) *Ontological relativity and other essays.* Columbia University Press. [aPJHS]

Rachlin, H. (1985) Maximization theory and Plato's concept of the good. *Behaviorism* 13(1):3–20. [aPJHS]

Rachlin, H. & Green, L. (1972) Commitment, choice and self-control. *Journal of the Experimental Analysis of Behavior* 17:15–22. [JEM]

Rachlin, H., Battalio, R., Kagel, J. & Green, L. (1981) Maximization theory in behavioral psychology. *Behavioral and Brain Sciences* 4:371–417. [JEM]

Ramsey, F. P. (1929/1990) General propositions and causality. In: *Philosophical papers,* ed. D. H. Mellor. Cambridge University Press. [N-ES]

Rapoport, A., Guyer, M. & Gondon, O. (1976) *The 2 × 2 game.* University of Michigan Press. [APW]

Rapport, D. J. (1971) An optimization model of food selection. *American Naturalist* 105:575–88. [aPJHS]

Rescher, N. (1979) *Leibniz: An introduction to his philosophy.* Basil Blackwell. [aPJHS]

Rodman, P. S. & Cant, J. G. H., eds. (1984) *Adaptations for foraging in nonhuman primates.* Columbia University Press. [aPJHS]

Roitblat, H. L. (1982) Decision making, evolution and cognition. In: *Evolution and determinism of animal and human behavior,* ed. H. D. Schmidt & G. Tembrock. Springer-Verlag. [DAH]

Roll, R. (1989) The international crash of October 1987. In: *Black Monday and the future of financial markets,* ed. R. W. Kamphuis, R. C. Kormendi & J. W. H. Watson. Dow Jones-Irwin, Inc. [aPJHS]

Rosen, R. (1967) *Optimality principles in biology.* Butterworths. [aPJHS]

Rosenberg, A. (1976) *Microeconomic laws.* Pittsburgh University Press. [aPJHS]

(1985) *The structure of biological science.* Cambridge University Press. [aPJHS]

Rosenblueth, A. Wiener, N. & Bigelow, J. (1943) Behavior, purpose and teleology. *Philosophy of Science* vol. 10(1):18–24. [aPJHS]

Rothschild, M. (1973) Models of market organization with imperfect information: A survey. *Journal of Political Economy* 81:1283–1308. [aPJHS]

(1974) Searching for the lowest price when the distribution of prices is unknown. *Journal of Political Economy* 82:689–711. [arPJHS]

Rumelhart, D. E., McClelland, J. L. & the PDP research group (1986) *Parallel distributed processing: Explorations in the microstructure of cognition.* II. Bradford Books. [JRA]

Russell, B. (1959) *Mysticism and logic.* George Allen & Unwin Ltd. [aPJHS, KMS]

Sahlin, N.-E. (1990) *The philosophy of F. P. Ramsey.* Cambridge University Press. [N-ES]

Salton, G. & McGill, M. J. (1983) Introduction to modern information retrieval. McGraw-Hill. [JRA]

Samuelson, P. (1946) *Foundations of economic analysis.* Harvard University Press. [aPJHS]

(1963) Discussion: Problems of methodology. *American Economic Review* 53(suppl.):227–36. [aPJHS]

(1970) Maximum principles in analytical economics. Nobel Memorial Lecture, Dec. 11. [aPJHS]

Satterthwaite, M. A. (1975) Strategy-proofness and Arrow's conditions. *Journal of Economic Theory* 1:187–217. [aPJHS]

Sayre, K. M. (1986) Intentionality and information processing: An alternative model for cognitive science. *Behavioral and Brain Sciences* 9:121–37. [aPJHS]

Schelling, T. C. (1960) *The strategy of conflict.* Harvard University Press. [aPJHS]

Schlick, M. (1939) When is man responsible? In: *Problems of ethics.* Prentice-Hall, Inc. [aPJHS]

Schoemaker, P. J. H. (1982) The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature* 20:529–63. [aPJHS, ESh]

(1984) Optimality principles in science: Some epistemological issues. In: *The quest for optimality,* ed. J. H. P. Paelinck & P. H. Vossen. Gower. [arPJHS]

(1990) Strategy, complexity and economic rent. *Management Science* 36(3). [aPJHS]

Schoemaker, P. & Hershey, J. (1991) Utility measurement: Signal, noise and bias. In: *Organizational Behavior and Human Decision Processes* (in press). [rPJHS]

Schroedinger, E. (1954) *Nature and the Greeks.* Cambridge University Press. [aPJHS]

(1967) *What is life?—Mind and matter.* Cambridge University Press (Cambridge). [aPJHS]

Scitovsky, T. (1976) *The joyless economy.* Oxford University Press. [rPJHS]

Shannon, C. E. & Weaver, W. (1949) *The mathematical theory of communications.* University of Illinois Press. [aPJHS]

Shepard, R. N. (1964) On subjectively optimum selections among multi-attribute alternatives. In: *Human judgment and optimality,* ed. M. W. Shelby & G. L. Bryan Wiley. [aPJHS]

(1987) Evolution of a mesh between principles of the mind and regularities of the world. In: *The latest of the best,* ed. J. Dupré. MIT Press. [aPJHS]

Shiller, R. (1981) Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 7:421–36. [aPJHS]

Simon, H. A. (1957) *Models of man: Social and rational.* Wiley. [arPJHS]

(1981) *The sciences of the artificial,* 2nd ed. MIT Press. [arPJHS]

(1983) *Models of bounded rationality.* MIT Press. [PM]

Skinner, B. F. (1969) *Contingencies of reinforcement.* Appleton-Century-Crofts. [aPJHS]

Slovic, P. & Lichtenstein, S. (1983) Preference reversals: A broader perspective. *American Economic Review* 73:596–605. [aPJHS]

Slovic, P. & Tversky, A. (1974) Who accepts Savage's axiom? *Behavioral Science* 19:368–73. [aPJHS]

Slovic, P., Lichtenstein, S. & Fischhoff, B. (1989) Decision making. In: *Steven's handbook of experimental psychology,* ed. R. C. Atkinson, R. J. Herrnstein, G. Lindzey & R. D. Luce. Wiley. [ESh]

Smith, A. (1776/1976) *The wealth of nations.* University of Chicago Press. [aPJHS]

Smith, E. A. (1987) Optimization theory in anthropology: Applications and critiques. In: *The latest on the best,* ed. J. Dupré. MIT Press. [aPJHS]

Sober, E. (1988) *Reconstructing the past: Parsimony, evolution, and inference.* MIT Press. [ESo]
(in press) Let's razor Occam's razor. In: *Explanation and its limits,* ed. D. Knowles. Royal Institute of Philosophy, suppl. vol. 27. Cambridge University Press. (Cambridge) [ESo]
Solomon, R. L., Kamin, L. J. & Wynne, L. C. (1953) Traumatic avoidance learning: The outcome of several extinction procedures with dogs. *Journal of Abnormal Social Psychology* 48:291–302. [DSL]
Sopron, (1984) Plural rationality and interactive decision processes. Proceedings of the conference. Springer-Verlag. [AAW]
Sorenson, A. (1978) Mathematical models in sociology. *Annual Review of Sociology* 9:31–56. [aPJHS]
Spranca, M., Minsk, E. & Baron, J. (in press) Omission and commission in judgment and choice. *Journal of Experimental Social Psychology.*
Staddon, J. E. R. (1983) *Adaptive behavior and learning.* Cambridge University Press. [KMS]
Stigler, G. (1961) The economics of information. *Journal of Political Economy* 69:213–25. [arPJHS, PM]
Stritter, E. (1977) File migration. (Unpublished Ph.D. thesis STAN-CS-77-594.) Stanford University. [JRA]
Svirezhev, Y. M. (1972) Optimum principles in population genetics. In: *Studies on theoretical genetics,* ed. V. A. Ratner. Academy of Sciences, USSR, Novosibirsk. (In Russian with an English summary.) [JFC]
Teller, R. (1980) *The pursuit of simplicity.* Pepperdine University Press. [aPJHS]
Tinbergen, J. (1928) Minimum problemen in de Natuurkunde en Economie. Ph.D. thesis, University of Leiden, The Netherlands. [aPJHS]
Tipler, P. A. (1969) *Foundations of modern physics.* Worth. [aPJHS]
Tsotsos, J. (1988) A "complexity-level" analysis of intermediate vision. *International Journal of Computer Vision* 1:303–20. [DAM]
Tversky, A. & Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185:1124–31. [aPJHS, DSL]
(1981) The framing of decisions and the rationality of choice. *Science* 211:453–58. [aPJHS, DSL]
(1982) Evidential impact of base rates. In: *Judgment under uncertainty: Heuristics and biases,* ed. D. Kahnemann, P. Slovic & A. Tversky. Cambridge University Press (Cambridge). [SS-F]
(1986) Rational choice and the framing of decisions. *Journal of Business* 59(4) Part 2:251–78. [ESh]
Van Gastel, M. A. J. J. & Paelinck, J. H. P. (in press) Hypergraph conflict analysis. *Economics Letters.* [JHPP]
Varela, F. J. (1979) *Principles of biological autonomy.* Elsevier North-Holland. [aPJHS]
Vaughan, W. & Miller, H. L. (1984) Optimization versus response-strength accounts of behavior. *Journal of the Experimental Analysis of Behavior* 42:337–48. [MD]
Vergis, A., Steiglitz, K. & Dickinson, B. (1986) The complexity of analog computation. *Mathematics and Computers in Simulation* 28:91–113. [DAM]
Voltaire, M. de (1759/1966) *Candide* (trans. R. M. Adams). W. W. Norton. [aPJHS]
Von Helmholtz, H. (1886) Ueber Die Physikalische Bedeutung des Princips der Kleinsten Wirkung. *Journal für Mathematik,* Bd. C. (2):18–28. [aPJHS]
Von Neumann, J. & Morgenstern, O. (1947) *Theory of games and economic behavior,* 2nd ed. Princeton University Press. [arPJHS, ESh]
(1953) *Theory of games and economic behavior.* Princeton University Press. [OL]
Wagner, H. M. (1969) *Principles of operations research.* Prentice-Hall. [OL]
Warren, H. C. (1916) A study of purpose. *The Journal of Philosophy, Psychology, and Scientific Methods* 13(1–3). [aPJHS]
Wasserman, P. D. (1989) *Neural computing: Theory and practice.* Van Nostrand Reinhold. [DAH]
Whitehead, A. N. (1920) *The concept of nature.* Cambridge University Press. [aPJHS]
Wiener, N. (1961) *Cybernetics,* 2nd ed. MIT Press. [aPJHS]
Wigner, E. P. (1967) The unreasonable effectiveness of mathematics in the natural sciences. In: *Symmetries and reflections: Scientific essays of Eugene P. Wigner.* MIT Press. [FLB]
Williams, G. C. (1966) *Adaptation and natural selection.* Princeton University Press. [MDal]
Wilson, E. O. (1975) *Sociology: The new synthesis.* Harvard University Press. [aPJHS]
Wimsatt, W. C. (1972) Teleology and the logical structure of function statements. *Studies in the History of Philosophy of Science* 3(1):1–79. [aPJHS]
Winter, S. (1975) Optimization and evolution in the theory of the firm. In: *Adaptive economic models,* ed. R. H. Day & T. Groves. Academic Press. [PM]
Woodfield, A. (1976) *Teleology.* Cambridge University Press. [aPJHS]
Wright, L. (1976) Teleological explanation: An etiological analysis of goals and functions. University of California Press. [aPJHS]
Wright, S. (1954) The interpretation of multivariate systems. In: *Statistics and mathematics in biology,* ed. O. Kempthorne, T. A. Bancroft, J. W. Gowen & J. L. Lush. Iowa State College Press. [FLB]
Zipf, G. K. (1949) *Human behavior and the principle of least effort.* Addison-Wesley Press. [aPJHS]
Zucker, S. W., Dobbins, A. & Iverson, L. (1989) Two stages of curve detection suggest two styles of visual computation. *Neural Computation* 1:68–81. [DAM]

ERRATUM

Paul Bloom

Department of Psychology, University of Arizona, Tucson, AZ 85721

In my Response to the target article, "Natural Language and Natural Selection," co-authored by Steven Pinker and Paul Bloom (*BBS* 13(4):769), the journal printed the following sentence, "Rather, different mutations are stored independently in different lineages, and recombination brings them together to form vast numbers of new combinations, in people, in their descendents," in error. The sentence should have read: "Rather, different mutations are stored independently in different lineages, and recombination brings them together to form vast numbers of new combinations, in parallel, in their descendents."