

The Latest on the Best

Essays on Evolution and Optimality

edited by John Dupré

Chapter 13

From Evolution to Behavior: Evolutionary
Psychology as the Missing Link

Leda Cosmides and John Tooby

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England
1987

From Evolution to Behavior: Evolutionary Psychology as the Missing Link

Leda Cosmides and John Tooby

Popular wisdom has it that arguments against new ideas in science typically pass through three characteristic stages, from

1. "It's not true," to
2. "Well, it may be true, but it's not important," to
3. "It's true and it's important, but it's not new—we knew it all along."

If the papers in this volume are any indication, then the application of evolutionary biology to the understanding of human behavior has entered the "It's true but not important" stage.

Yet evolutionary theory is important for understanding human behavior, and not everyone knows it—in fact, those most involved in the scientific investigation of "human nature" are generally the most unaware of its implications. We shall argue that the reluctance of many social scientists to appreciate or take advantage of the richness of the evolutionary approach is a direct consequence of a widespread tendency to overlook a crucial link in the causal chain from evolution to behavior: the level of innate psychological mechanisms, described as information processing systems. This level is pivotal, because it describes the mechanisms that actually link the evolutionary process to manifest behavior. It is these mechanisms that evolve over generations; within any single generation it is these mechanisms that, in interaction with environmental input, generate manifest behavior. The causal link between evolution and behavior is made through the psychological mechanism.

Efforts that skip this step in the evolutionary analysis of behavior, as valuable as they may be in other ways, have contributed to an erroneous caricature of the evolutionary approach to behavior as offering nothing more than post hoc compilations of correspondences between behavior and loosely reinterpreted evolutionary theory. But a rejection of the evolutionary approach based on such an incomplete and misleading character-

We would like to thank David Buss, George Cosmides, and Roger Shepard for their very helpful comments on an earlier version of this paper, and Roger Shepard for all the encouragement and support he has given us.

ization of its nature and valid possibilities is mistaken: as we shall discuss, the search for order in human behavior requires the application of the emerging principles of evolutionary psychology. We shall argue that an approach drawn from evolutionary psychology, consistently applied, can repair many of the deficiencies that have hampered progress in the social sciences.

1 Natural Selection Theory Does Not Predict Invariance in the Manifest Behavior of Different Individuals

Sciences prosper when researchers discover the level of analysis appropriate for describing and investigating their particular subject: when researchers discover the level where invariance emerges, the level of underlying order. What is confusion, noise, or random variation at one level resolves itself into systematic patterns upon the discovery of the level of analysis suited to the phenomena under study. The lack of success the behavioral sciences have had since their founding has been explained either by the claim that no such science is possible (e.g., human complexity intrinsically transcends any attempt to discover fundamental patterns) or by the view we share, that progress has been slow because scientific efforts have not yet, for the most part, been framed using concepts and organizing principles suitable to the phenomena under study. Can such an appropriate level of inquiry be found for a science of human behavior? Because humans are the product of the evolutionary process, the explanation for their characteristics must be sought in the evolutionary process: for a science of human behavior, the level of underlying order is to be sought in an evolutionary approach.

However, using evolution as an informing concept is not enough. During the formative period of modern behavioral ecology in the 1970s, many researchers thought that evolutionary biology would revolutionize research in human behavior; this conviction spread after the publication of E. O. Wilson's *Sociobiology* drew widespread attention to the dramatic advances that were taking place in the application of evolution to behavior. Many thought that evolutionary theory would reveal the level of underlying order, that the apparent variation in human behavior would resolve itself into systematic patterns, that invariant relationships would be identified, and that a true social science would emerge. However, after more than a decade, this is a revolution still waiting to happen.

We shall argue that the reason that progress has been slow is that, in the rush to apply evolutionary insights to a science of human behavior, many researchers have made a conceptual "wrong turn," leaving a gap in the evolutionary approach that has limited its effectiveness. This wrong turn has consisted of attempting to apply evolutionary theory directly to the

level of manifest behavior, rather than using it as a heuristic guide for the discovery of innate psychological mechanisms.

The attempt to find evolutionary invariants at the level of manifest behavior has created a series of difficulties, from forced typological approaches, to using the "optimality" of manifest behavior (or the lack of it) as the measure of the success of the evolutionary paradigm. The assumption that manifest behavior should be invariant across individuals has invited a brute force, typological approach to variation in, for example, cross-cultural studies and primate behavior. All too often, the researcher would take the observed variation, average it, and typify the species or group by that average (see Tooby and DeVore, 1987, for a more extensive discussion of this problem). The variation itself is considered noise, or an embarrassment to be explained away. Those social scientists skeptical that biology had anything to offer to an understanding of human behavior would dwell on the extraordinary complexity of human behavior, and its enormous and engaging variety, and counterpose this richness to the clear explanatory inadequacy of what they considered to be naive and simplistic typological characterizations.

Yet natural selection theory itself predicts that the manifest behavior of different individuals will vary enormously. Furthermore, it deductively implies that an individual's behavior will often appear far from "optimal," when optimality is defined without respect to the individual's social environment. The reasons why this is so are summarized by Tooby and DeVore (1987), in their discussion of hominid behavioral evolution. They include the following:

1. The interests of different individuals are often in conflict; in fact, much of modern evolutionary theory analyzes the conflicting fitness interests of different categories of individuals [e.g., self versus kin (Hamilton, 1964), parent versus offspring (Trivers, 1974), male versus female (Trivers, 1972)]. An interaction between individuals whose fitness interests conflict cannot, in principle, produce an outcome that is optimal for both individuals. The outcome will either be optimal for one party but not the other, or it will be nonoptimal for both.
2. Therefore, larger patterns of social behavior are not necessarily optimal for any individual or group of individuals, but rather may be the emergent result of the conflicting interests of interacting individuals, each selected to promote its own inclusive fitness. Frequently, therefore, the behavior of an individual cannot be understood in isolation; its behavior will be the mutual result of its interests and the counterstrategies of those with whom the individual is associated.
3. Individuals are selected to be adapted to their individual situation, not simply to their local habitat. For example, an individual's best

behavioral strategy may depend on its size, its health, its aggressive formidability, its facility at accruing resources, or the number of sibs it can rely on for support. This means that organisms may be selected to be facultative strategists (where appropriate) rather than inflexibly committed to the same behavior or morphology. Consequently, individuals with the same psychological programming may manifest different behaviors in response to the different information they derive from assessing their own abilities and resources.

4. For certain social and reproductive behaviors, the favored strategy will depend on the distribution of other behaviors in the population [the prevailing analytic tool for dealing with this is game theory and evolutionarily stable strategies (Maynard Smith and Price, 1973)]. In such situations, selection can produce facultative psychological mechanisms that are sensitive to information indicating the distribution of relevant behaviors in the local population.

5. To be selected for, a trait need not be advantageous under every conceivable circumstance. It need only be of benefit *on balance*. This means it must be advantageous more often than not, or that the frequency with which it is advantageous, times the magnitude of the advantage, outweighs the frequency of disadvantage times the cost. Thus, selection for a trait is always against a background probability distribution of ancestral environmental conditions, and cannot be understood when abstracted from this background.

6. Therefore, natural selection cannot be expected to produce behavioral responses that maximize fitness under every imaginable circumstance. The situational specificity of adaptation depends on the selective history of similar situations. The degree of situational adaptation manifested by individuals will be a matter of (a) how common in the species' evolutionary history that situation has been, (b) how long (in phylogenetic terms) it has been recurring, and (c) how large its fitness consequences are. Organisms will be well adapted to common, important situations, reasonably adapted to common less important situations and uncommon highly important situations, but not adapted to uncommon, unimportant situations.

7. The recognition that adaptive specializations have been shaped by the statistical features of ancestral environments is especially important in the study of human behavior. Our species spent over 99% of its evolutionary history as hunter-gatherers in Pleistocene environments. Human psychological mechanisms should be adapted to those environments, not necessarily to the twentieth-century industrialized world. The rapid technological and cultural changes of the last several thousand years have created many situations, both important and unimportant, that would have been uncommon (or nonexistent) in

Pleistocene conditions. Evolutionary theorists ought not to be surprised when evolutionarily unprecedented environmental inputs yield maladaptive behavior. Our ability to walk fails us hopelessly when we are chased off a cliff.

Consequently, behavioral variation is not an embarrassment to evolutionary theory, it is a prediction of evolutionary theory. Equally, the assumption that individuals pursue strategies that will tend to promote their inclusive fitness deductively entails that (1) an individual's theoretically "optimal" behavioral strategy will vary, depending on the composition of its social group, and (2) an interaction between individuals whose fitness interests conflict cannot, in principle, produce an outcome that is optimal for both individuals. Typological approaches to manifest human behavior, involving attempts to interpret such behavior in terms of evolutionary optimality, violate these deductive implications of natural selection theory. For these and other reasons, the search for invariance on the level of manifest behavior will have very limited success.

When the appropriate level of analysis is found, variation becomes fuel in the search for order: instead of averaging out variation, one looks for systematic relations among the different varying elements. What is variable at one level manifests order—that is, invariance—at another. Instead of lamenting the complex variations in human behavior, researchers can use patterns in behavioral variation positively, as clues to the nature of the psychological mechanisms that produce behavior.

2 *Evolution → Psychological Mechanism → Behavior*

To speak of natural selection as selecting for "behaviors" is a convenient shorthand, but it is misleading usage. The error is worth belaboring, because the failure to appreciate it has delayed the fruitful application of evolutionary theory to human behavior by years. When used too casually, this shorthand misleads because it obscures the most important level of proximate causation: the psychological mechanism.

Natural selection cannot select for behavior per se; it can only select for mechanisms that produce behavior. There is nothing special about behavior in this regard; the same can be said, for example, of digestion. Natural selection can only rearrange patterns in tissues and molecules; these rearrangements have *effects*, and it is because they have these effects that they are selected for or not. Natural selection gives us teeth, salivary amylase, a peristaltic esophagus, an acid-filled stomach, an absorptive colon: mechanisms that produce digestion. The operation of these mechanisms causes certain molecules to be extracted from plant and animal tissues and incorporated into our own tissues: an effect that we call digestion. Natural selection gives us food processing machinery, and the operation of this

machinery results in digestion, which is an effect of the functioning of mechanisms.

Behavior, like digestion, is an effect of the functioning of mechanisms. Natural selection can give you a reflex arc, and the functioning of this arc causes an effect: your leg swings when your knee is tapped. But this effect cannot occur in the absence of a mechanism for producing it. Behavior cannot occur *sui generis*; behavior is an effect produced by a causal system: proximately, by psychological mechanisms. Although researchers would acknowledge these points as patently obvious, in practice, many simply methodologically leapfrog this level, with unfortunate consequences such as those discussed. Their desire to do this stems, in many cases, from the belief that the exploration of mechanisms means the exploration of the neurophysiological bases of behavior, a difficult endeavor, and one that, at the present state of knowledge, is limited to addressing only very simple kinds of behaviors. However, there exists an alternative approach to the study of psychological mechanisms that does not involve neurophysiology, with its present limitations. This is the characterization of psychological mechanisms in terms of their information processing structure. This approach dovetails smoothly with evolution, because in the adaptive regulation of behavior, information is key.

Behavior is not randomly emitted; it is elicited by information, which is gleaned from the organism's external environment, and, proprioceptively, from its internal states. Natural selection gave us information processing machinery to produce behavior, just as it gave us food processing machinery to produce digestion. This machinery selects—and frequently seeks—particular information from the environment; it manipulates it, extracts inferences from it, stores some of it in memory in altered form; the machinery's output is used to make mental models, to inform other parts of the system, and to instruct the motor neurons responsible for behavior. *The evolutionary function of the human brain is to process information in ways that lead to adaptive behavior*; the mind is a description of the operation of a brain that maps informational input onto behavioral output.

Thus, behavior is one output of our information processing machinery. Behavioral output differs with informational input; the information processing machinery that maps informational input onto behavioral output is a psychological mechanism.

The psychology of an organism consists of the total set of proximate mechanisms that cause behavior. Natural selection, acting over evolutionary time, shapes these mechanisms so that the behavior of the organism correlates to some degree with its fitness. However, in the lifetime of any particular animal, it is the proximate mechanisms that actually cause behavior—not natural selection. If these proximate mechanisms can be understood, behavior can be predicted more exactly; understanding the

fitness-promoting strategies studied by evolutionary theorists allows only approximate prediction. Behavior correlates exactly with proximate mechanisms, but only approximately with the fitness-promoting strategies that shaped those mechanisms.

Evolutionary psychology (Tooby, 1985) relates explanations in terms of adaptive strategy to explanations in terms of proximate mechanisms. Correct characterization of adaptive strategies gives precise meaning to the concept of function for proximate mechanisms. Reciprocally, a detailed analysis of the proximate mechanisms of a species gives rich insight into the present and past selective pressures that have acted on it. Psychological mechanisms constitute the missing causal link between evolutionary theory and behavior. Evolutionary theory frequently appears to lack predictive value because most researchers skip this crucial predictive and explanatory level. Yet it is the proximate mechanisms that cause behavior that promise to reveal the level of underlying order for a science of human behavior.

3 *The Cognitive Level of Explanation*

Psychological mechanisms can be studied on different descriptive and explanatory levels. Most biologically informed studies of proximate mechanisms have described psychological mechanisms in terms of their physiological underpinnings, finding, for example, that birth spacing is mediated by lactation, which suppresses ovulation, that testosterone levels change with shifts in dominance, thereby affecting agonistic behavior, or that one part of the brain controls language while another part controls sexual behavior.

But natural selection theory, so far, has made only limited contributions to the investigation of physiology. Just as different kinds of hardware can run the same computer program, different physiological mechanisms can accomplish the same adaptive function. Both humans and pitcher plants digest animal tissues, but the physiological mechanisms by which humans and pitcher plants accomplish this function are different. And there is another, pragmatic problem: unless you know that a particular information processing system exists and what its function is, it is very difficult to discover its physiological underpinnings. Who would look for the physiological mechanisms responsible for the contraction of the heart unless they first knew that the heart exists and that its function is to pump blood?

Although valuable, physiological studies do not address a crucial *functional* level of explanation, a level that describes what a mechanism does, rather than how it does it. Evolutionarily oriented students of human behavior have neglected what may prove to be the most important level of proximate causation: the cognitive level. *Adaptive behavior is predicated on adaptive thought*: an animal must process information from its environment in ways that lead to fit behaviors while excluding unfit behaviors. The

cognitive level of explanation describes psychological mechanisms in functional terms, as programs that process information.¹

Traditionally, ethologists have studied very simple cognitive programs: a newborn herring gull has a cognitive program that defines a red dot on the end of a beak as salient information from the environment, and that causes the newborn to peck at the red dot upon perceiving it. Its mother has a cognitive program that defines pecking at her red dot as salient information from her environment, and that causes her to regurgitate food into the newborn's mouth when she perceives its pecks.

Note that the descriptions of these simple programs are entirely in terms of the functional relationships among different pieces of information; they describe two simple information processing systems. Naturally, these programs are instantiated in some kind of neurological "hardware." However, knowledge of this hardware would add little to our understanding of these programs as information processing systems—presumably, one could build a silicon-based robot that would produce the same behavioral output in response to the same informational input. The robot's cognitive programs would maintain the same functional relationships among pieces of information, and therefore be identical to the cognitive programs of the herring gull. However, the robot's "neural" hardware would be totally different. The specification of a cognitive program constitutes a complete description of an important level of proximate causation, independent of any knowledge of the physiological mechanisms by which the program is instantiated.

We assume that the cognitive programs of different individuals² of a species are essentially the same—that cognitive programs are species-typical traits. However, the parameters fed into them can be expected to differ with individual circumstance. Insofar as individual variation in personal qualities (such as aggressive formidability or sexual attractiveness), in opportunities to engage in particular behaviors (to mate, to threaten, to help), and in the social and physical environment, are all parameters that feed into the same cognitive programs, variations in these parameters will produce variations in manifest behavior across individuals. Therefore, although the cognitive programs of different individuals should be essentially the same, the manifest behavior of different individuals may be different.³ Cognitive programs constitute the level of invariance for a science of human behavior, not behavior itself.

When applied to behavior, natural selection theory is more closely allied with the cognitive level of explanation than with any other level of proximate causation. This is because the cognitive level seeks to specify a psychological mechanism's function, and natural selection theory is a theory of function. Natural selection theory specifies how an organism should respond to different kinds of information from its environment. It defines adaptive information process-

ing problems that the organism must have some means of solving. Cognitive programs are solutions to information processing problems.

An evolutionary approach to understanding the cognitive level of proximate causation asks, What kind of programming must an organism have if it is to extract and process information about its environment in a way that will lead to adaptive behavior? How does the organism use information from its environment to compute what constitutes the "right" behavior at the right place and the right time (Staddon, this volume)?

4 Evolution and the Cognitive Level

It is nearly impossible to discover how a psychological mechanism processes information unless one knows what its function is, what it was "designed" or selected to do. Trying to map out a cognitive program without knowing what its function is, is like attempting to understand a computer program by examining it in machine language, without knowing whether it is for editing text, accounting, or launching the Space Shuttle. It is possible that a gifted programmer may finally figure it out, but not probable. If, on the other hand, the programmer knows that the program she is trying to map out is a text editor, she can begin by looking for a way of loading text, or for a command that will delete a word, or for a procedure that will move a whole paragraph. It is far easier to understand the architecture of a "black box" if one knows what it was designed to do.

Recognizing this, a number of cognitive scientists, such as Chomsky, Shepard, Fodor, and Marr, recently have argued that the best way to understand any mechanism, either mental or physical, is first to ask what its purpose is, what problem was it designed to solve (e.g., Chomsky, 1975; Shepard, 1981; Fodor, 1983; Marr and Nishihara, 1978).

This is exactly the question that evolutionary theory allows one to address—it allows one to pinpoint the kinds of problems the human mind was "designed" to solve, and consequently should be very good at solving. And although it cannot tell one the exact structure of the cognitive programs that solve these problems, it can suggest what design features they are likely to have. It allows one to develop a "computational theory" for that problem domain: a theory specifying what functional characteristics a mechanism capable of solving that problem must have (Marr, 1982; Marr and Nishihara, 1978).

Many cognitive psychologists assume that the human mind is a general-purpose computer with domain-general, content-independent processes. We shall argue that from an evolutionary point of view, this is a highly implausible and unparsimonious assumption, and logically impossible to sustain. There are domains of human activity for which the evolutionarily appropriate information processing strategy is complex, and deviations

from this strategy result in large fitness costs. An organism that relied on the vagaries of trial-and-error learning for such domains would be at a selective disadvantage (see also Shepard, 1981).

Instead, for such domains, humans should have evolved "Darwinian algorithms"—specialized learning mechanisms that organize experience into adaptively meaningful schemas or frames (Cosmides, 1985). When activated by appropriate environmental or proprioceptive information, these innately specified "frame-builders" should focus attention, organize perception and memory, and call up specialized procedural knowledge that will lead to domain-appropriate inferences, judgments, and choices. Like Chomsky's language acquisition device, these inference procedures allow you to "go beyond the information given"—to reason adaptively even in the face of incomplete or degraded information (Bruner, 1973).

There are many domains of human activity that should have Darwinian algorithms associated with them. Aggressive threat, mate choice, sexual behavior, pair-bonding, parenting, parent-offspring conflict, friendship, kinship, resource accrual, resource distribution, disease avoidance, predator avoidance, and social exchange are but a few. The dynamics of natural selection rigidly constrain the patterns of behavior that can evolve in such domains, and therefore provide insights into the structure of the cognitive programs that produce these patterns.

In the remainder of this article we present arguments supporting this perspective.

5 Complex Adaptive Problems Should Be Defined in Computational Theories

The signal lesson lurking beneath the surface of modern evolutionary theory is that adaptive behavior requires the solution of many information processing problems that are highly complex—far more complex than commonly supposed. The cognitive programs that allow the newborn herring gull to gain sustenance from its mother are relatively simple: they directly connect the perception of an environmental cue with an adaptively appropriate behavioral response. But not all adaptive problems are so easily solved, and many complex adaptive problems can be solved only by complex cognitive programs.

Discovering the structure of complex cognitive programs requires a great deal of theoretical guidance. A series of hunt-and-peck experiments may uncover a few simple cognitive programs, but it is unlikely that a research program that is blind to function will ever uncover the structure of a complex information processing system—such as the human mind.

What form should this theoretical guidance take? In his pioneering studies of visual perception, David Marr argued that "computational theories" of each information processing problem must be developed

before progress can be made in experimentally investigating the cognitive programs that solve them (e.g., Marr, 1982; Marr and Nishihara, 1978). A computational theory specifies the nature of an information processing problem. It does this by incorporating "constraints on the way the world is structured—constraints that provide sufficient information to allow the processing to succeed" (Marr and Nishihara, 1978, p. 41). A computational theory is an answer to the question, What must happen if a particular function is to be accomplished?

For example, the information processing problem that Marr wanted to understand was how an organism reconstructs three-dimensional objects in the world from a two-dimensional retinal display. As you walk around a table with a square top, for example, light reflected from the tabletop hits your retina, projecting upon it a two-dimensional trapezoid of changing dimensions. Yet you do not perceive an ever-deforming, two-dimensional trapezoid. Instead, your cognitive programs use these data to construct a "percept" of a stable, three-dimensional, square tabletop.

To understand how we compute solid objects from data like this, Marr and his colleagues first examined relevant constraints and relationships that exist in the world, like the reflectant properties of surfaces. They consider the discovery of such constraints the "critical act" in formulating a theory of this computation, because these constraints must somehow be used by and embodied in any cognitive mechanism capable of solving this problem (Marr, 1982; Marr and Nishihara, 1978). Marr calls the specification of such constraints, together with their deductive implications, a "computational theory" of an information processing problem.

Natural selection, in a particular ecological situation, defines and constitutes "valid constraints on the way the world is structured," and therefore can be used to create computational theories of adaptive information processing problems. For example, the cognitive programs of an organism that confers benefits on kin cannot violate the [cost to self < (benefit to kin member) × (coefficient of relatedness to kin member)] constraint of kin selection theory. Cognitive programs that violate this constraint cannot be selected for. Cognitive programs that instantiate this constraint can be selected for. This is inherent in the dynamics of natural selection, true of any species on any planet at any time. A species may lack the ability to confer benefits on kin, but if it has such an ability, then it has it by virtue of cognitive programs that produce behavior that respects this constraint.

The production of behavior that respects constraints imposed by the evolutionary process is a cognitive program's *adaptive function*: the reason it was selected for, the reason it could outcompete other cognitive programs and spread through the population to become a species-typical trait.

The specification of constraints imposed by the evolutionary process—

the specification of an adaptive function—does not, in itself, constitute a complete computational theory. These constraints merely define what counts as adaptive behavior. Cognitive programs are the means by which behavior—adaptive or otherwise—is produced. The important question for a computational theory to address is, What kind of cognitive programs must an organism have if it is to behave adaptively?

Natural selection theorists do not usually think of their theories as defining information processing problems, yet this is precisely what they do. For example, kin selection theory raises—and answers—questions such as, How should the information that *X* is your brother affect your decision to help him? How should your assessment of the cost to you of helping your brother, versus the benefit to your brother of receiving your help, affect your decision? Will the information that *Y* is your cousin have a different effect on your decision than if you thought *Y* were your brother? In general, how should information about your relatedness to *X*, the costs and benefits to you of what *X* wants you to do for him, and the costs and benefits to *X* of your coming to his aid, affect your decision to help *X*?

As these questions show, an organism's behavior cannot fall within the bounds of the constraints imposed by the evolutionary process unless it is guided by cognitive programs that can solve certain information processing problems that are very specific. To confer benefits on kin in accordance with the constraints of kin selection theory, the organism must have cognitive programs that allow it to extract certain specific information from its environment: who are its relatives? which kin are close and which distant? what are the costs and benefits of an action to itself? to its kin? The organism's behavior will be random with respect to the constraints of kin selection theory unless (1) it has some means of extracting information relevant to these questions from its environment, and (2) it has well-defined decision rules that use this information in ways that instantiate the theory's constraints. A cognitive system can generate adaptive behavior only if it can perform specific information processing tasks such as these.

The fact that any organism capable of conferring benefits on its kin must have cognitive programs capable of solving these information processing problems does not imply that different species will solve each problem via the same cognitive program. There are many reasons why such programs may differ. For example, different environmental cues may have different reliabilities and accessibilities for different species. Moreover, each species occupies a different ecological niche, and hence the value of particular actions will differ across species: the cognitive programs of a baboon will assign a different value to social grooming than will the cognitive programs of a whale. But cognitive programs that perform the same function in different species may differ in more profound ways: the cognitive programs for recognizing kin might operate through phenotype matching

in one species, but through early imprinting in another species. Both programs will accomplish the same important adaptive function. Yet they will embody radically different information processing procedures, and they will process different information from the environment.

Natural selection theory can be used to develop computational theories of adaptive information processing problems. As we shall show below, such computational theories are valuable as heuristic guides for psychological research, despite the fact that evolutionary theory does not uniquely specify which cognitive programs will be used to accomplish a given function.

6 *The Importance of Computational Theories*

The most essential part of a computational theory is a catalog of the specific information processing problems entailed by the constraints of natural selection theory. They should be made explicit, for they are the building blocks of psychological theories. There are two reasons why this is so.

The first is obvious. Knowing, for example, that an organism must have some means of distinguishing kin from nonkin may not uniquely determine the structure of a cognitive program, but it does help narrow hypotheses. The cognitive program responsible must be sensitive to environmental cues that correlate with kin, but do not correlate with non-kin. In most cases, very few cues from the species environment of evolutionary adaptedness will be sufficiently reliable or accessible, and the researcher can very quickly discover which are used by the organism's cognitive programs. Discovering which cues are used will illuminate other of the program's information processing procedures: early exposure suggests an imprinting process, whereas facial similarity suggests phenotype matching procedures. Step by step, deduction by deduction, the cognitive programs responsible for kin recognition can be mapped. In the meantime, the researcher who is blind to function will not even be looking for a program that guides kin recognition, let alone figure out which environmental stimuli it monitors, and how it processes them.

The second reason why a fully elaborated computational theory is essential is less obvious, but far more important. The computational theory allows a test of adequacy that any proposed psychological theory must be able to pass. The test is this: *Is the hypothesized system of cognitive programs powerful enough to realize the computational theory? That is, is the proposed mechanism capable of solving the adaptive problem?*

Any proposed cognitive system must be powerful enough to produce adaptive behaviors while *not* simultaneously producing maladaptive behaviors. Not just any cognitive program will do: our cognitive programs

must be constructed in such a way that they somehow lead to the adaptive results specified by evolutionary theory on the basis of the information available. This crucial test of adequacy may allow researchers to eliminate whole categories of hypotheses, for current research in cognitive psychology and artificial intelligence suggests that many of the general-purpose learning theories that were popular in psychology's past are not powerful enough to solve even simple computational problems, let alone the complex problems posed by natural selection theory.

7 The Computational Theory Test

Thirty years ago, the study of the psychology of language took a major stride forward when Noam Chomsky developed a computational theory that allowed him to test whether certain hypothesized learning mechanisms were powerful enough to account for how humans acquire the ability to produce grammatical sentences. By this method, he was able to falsify the hypothesis that humans learn language through operant conditioning. Subsequently, others have used this method as a primary tool in constructing alternative psychological theories of language that are more powerful, and therefore more promising (for review, see Wanner and Gleitman, 1982). This incident shows that the "computational theory test" can provide an enormously effective tool for psychological theory.

Chomsky's (1957, 1959) computational theory was the grammar of the English language: a set of rules that can generate all the grammatical sentences of English, but no ungrammatical sentences. The information processing problem to be solved was, How do we learn this grammar? Can it be learned via the simple, stimulus-response (S-R) information processing mechanisms proposed by the behaviorists of the time, or does the acquisition of a natural language grammar require cognitive programming that is more specialized and complex?

Chomsky demonstrated that the general-purpose, S-R learning mechanisms proposed by the behaviorists were not powerful enough to allow one to acquire English grammar: they were not powerful enough to permit the speaker to produce many grammatical sentences, nor could they prevent the speaker from producing many ungrammatical sentences.

Native speakers of English have internalized its grammar; Chomsky showed that the behaviorists' learning mechanisms could not, in principle, account for this fact. He thereby falsified the hypothesis that we acquire grammar via such mechanisms. In fact, this computational theory test allowed him to eliminate a whole class of hypotheses: those invoking learning mechanisms that embody a "finite state grammar" (Chomsky, 1957).

This demonstration was an important turning point in the development

of modern psychology. Up until that point, psychology had been dominated by behaviorism's general-purpose learning theories. These theories were *domain general*: the same process was supposed to account for learning in all domains of human activity, from suckling at the breast to the most esoteric feat of modern technology. Yet by specifying what actually needed to be accomplished in order to produce grammatical utterances, Chomsky showed that a task routinely mastered by two-year-old children was too complexly structured to be accounted for by behaviorist learning theory.

Chomsky's specification of a computational theory convinced many psychologists that no general-purpose learning mechanism would be powerful enough to permit the acquisition of the grammar of a natural language under natural conditions. But what kind of learning mechanism *would* have the requisite power? Chomsky (1980) argued that just as the body has many different organs, each of which is specialized for performing a different function—a heart for pumping blood, a liver for detoxifying poisons—the mind can be expected to include many different “mental organs.” A mental organ is an information processing system that is specialized for performing a specific cognitive function. A mental organ instantiates learning theories that are *domain specific*: its procedures are specialized for quick and efficient learning about an evolutionarily important domain of human activity. Chomsky argued that the acquisition of a grammar could be accomplished only through a highly structured and complex “language acquisition device”: a functionally distinct mental organ that is specialized for learning a language.

The controversy between Chomsky and the behaviorists has broad applicability. Many psychologists think of it as a controversy about innateness, but, as we shall see below, it was not. “Innate” is not the “opposite” of “learned.” Every coherent learning theory—even Hume's associationism—assumes the existence of innate cognitive mechanisms that structure experience.⁴ A “blank slate” will stay forever blank: Without innate cognitive mechanisms, learning is impossible (e.g., Hume, 1977/1748; Kant, 1966/1781; Quine, 1969; Popper, 1972). Rather, the controversy in psycholinguistics is important because it highlights the ambiguity of the most central concept in the history of psychology: learning.

8 “Learning” Is Not an “Alternative Hypothesis”

Many common concepts in the social sciences are used as if they are hypotheses and explanations, but in fact are not. “Learning” is a concept that many people believe is fully freighted with meaning; analytically, however, the only meaning to the word “learned” is “environmentally

influenced." As a hypothesis to account for mental or behavioral phenomena, it is nearly devoid of meaning. Processes categorized as "learning" are accomplished through information processing mechanisms. Such mechanisms may be simple or complex, domain general, or domain specific. An organism may have many different learning mechanisms, or just a few. The belief that the human mind contains only one, simple, domain general cognitive process that results in "learning"—be it "induction" or "hypothesis testing" or "conditioning"—is nothing but conjecture. It has no basis in fact, and can only be explained as a metatheoretical holdover from the heyday of behaviorism.

In reality, the controversy in psycholinguistics was over whether the innate learning mechanisms that allow humans to acquire a grammar are simple and domain general or complex and domain specific (e.g., Atherton and Schwartz, 1974; Chomsky, 1975; Katz, 1975; Marshall, 1981; Putnam, 1967). The behaviorists thought that the simple, domain general processes of classical and operant conditioning could account for language; Chomsky showed that they could not, and proposed the existence of learning mechanisms that were complex and domain specific. Both camps agreed that language is "learned"; they disagreed about *how* it is learned.

The failure to grasp this point leads to enormous conceptual confusion in the behavioral sciences. The common belief that "learning" is an alternative hypothesis to an evolutionary theory of adaptive function is a category error. Learning is a cognitive process. An adaptive function is not a cognitive process; it is a problem that is solved by a cognitive process. Learning is accomplished through psychological mechanisms (whose nature is not yet understood), and these were created through the evolutionary process, which includes natural selection. Consequently, the issue is not whether a behavior is the result of natural selection "or" learning. The issue is, What kind of learning mechanisms would natural selection have produced?

When models of cognitive programs become sufficiently well specified actually to account for empirical results, they often turn out to be complex and domain specific. When researchers present such well-specified models together with the empirical results that support them, they are often met with the counter-claim that "people might learn to think that way." Yet, the invocation of an unspecified learning process does not constitute a valid alternative hypothesis. Suggesting that "learning" is an alternative hypothesis is comparable to claiming that an alternative hypothesis to a well-specified theory of vision, such as Marr's (1982), is, "Light hits the retina and this causes the organism to see three-dimensional objects." This is not an explanation; it is a description of the phenomenon to be explained. All the intervening steps are missing; it does not count as an "alternative hypothesis" because no one has bothered to specify the nature of the cognitive programs that cause it to happen.

"Learning" designates the phenomenon to be explained. A complex, domain specific cognitive program *is* a learning mechanism; how, then, can "learning" be construed as an "alternative hypothesis"?

The claim that a behavior is the product of "culture" is not an "alternative hypothesis" either. It entails nothing more than the claim that surrounding or preceding individuals are an environmental factor that have influenced the behavior under discussion in some way. It leaves the learning mechanisms that allow humans to acquire and generate culture completely unspecified (Tooby and Cosmides, 1987).

In speaking with evolutionary biologists and evolutionarily oriented anthropologists, we find that many operate from the implicit premise that an organism can "decide" which course of action, however complex, will maximize its inclusive fitness simply by inspecting the environment. These researchers interpret the fact that humans were produced by the evolutionary process to mean that humans must be maximizing their inclusive fitness in all situations—even in evolutionarily unprecedented modern environments. This view makes sense only if one believes that the organism has a "simple" cognitive program that says, "Do that which maximizes your inclusive fitness." Yet this is merely a veiled way of claiming that the organism "learns" what to do to maximize its fitness. It is not a hypothesis. It leaves "learning" a mysterious, omniscient, and utterly unspecified process.

It is improper to invoke an undefined process as an explanation. "Learning" should not be invoked to explain other phenomena at this point in the history of science, because it is itself a phenomenon that requires explanation. The nature of the cognitive processes that allow learning to occur are far from understood.

The tendency to assume that learning is accomplished only through a few simple domain general mechanisms lingers in cognitive psychology. We believe this metatheoretical stance is seriously flawed, and persists only because psychologists and evolutionary biologists have not joined forces to create computational theories that catalog the specific and detailed information processing problems entailed by the need to track fitness under Pleistocene conditions. Below, we join Rozin (1976), Shepard (1981), and Symons (1987) in arguing that a consideration of such problems suggests that natural selection has produced a great many cognitive programs that are complex and highly domain specific.

In this article, cognitive programs that evolved to accomplish important adaptive functions are called "Darwinian algorithms" (Cosmides, 1985).⁵ We now turn to the question, Does natural selection theory suggest that most Darwinian algorithms will be domain general, or domain specific?

9 *Why Should Darwinian Algorithms Be Specialized and Domain Specific?*

Nature has kept us at a great distance from all her secrets, and has afforded us only the knowledge of a few superficial qualities of objects; while she conceals from us those powers and principles, on which the influence of these objects entirely depends. Our senses inform us of the colour, weight, and consistence of bread; but neither sense nor reason can ever inform us of those qualities, which fit it for the nourishment and support of a human body. (David Hume, 1977/1748, p. 21)

Genes coding for psychological mechanisms that promote the inclusive fitness of their bearers will outcompete those that do not, and tend to become fixed in the population. The promotion of inclusive fitness is an evolutionary "end"; a psychological mechanism is a means by which that end is achieved. Can the human mind be comprised primarily of domain general and content-independent psychological mechanisms, and yet realize this evolutionary end? We shall argue that natural selection could not have produced such a psyche, nor could such a hypothetical psyche successfully promote fitness, that is, regulate behavior adaptively.

Consider how Jesus explains the derivation of the Mosaic code to his disciples:

Jesus said unto him, "Thou shalt love the Lord, thy God, with all thy heart, and with all thy soul, and with all thy mind. This is the first and great commandment. And the second is like it, Thou shalt love thy neighbor as thyself. *On these two commandments hang all the law and the prophets.*" (Matthew 22 : 37–40, emphasis added)

Jesus has given his disciples a domain general, content-independent decision rule to be used in guiding their behavior. But what does it mean *in practice*? Real life consists of concrete, specific situations. How, from this rule, do I infer what counts as "loving my neighbor as myself" when, to pick a standard biblical example, my neighbor's ox falls into my pit? Should I recompense him, or him me? By how much? How should I behave when I find my neighbor sleeping with my spouse? Should I fast on holy days? Should I work on the Sabbath? What counts as fulfilling these commandments? How do I know when I have fulfilled them?

In what sense does all the law "hang" from these two commandments?

These derivations are not obvious or straightforward. That is why the Talmud was written. The Talmud is a "domain specific" document: an interpretation of the "law" that tells you what actions fulfill the injunctions to "love God" and "love your neighbor" in the concrete, specific situations you are likely to encounter in real life. The Talmud solves the "frame

problem" (e.g., Boden, 1977; Fodor, 1983) posed by a "domain general" rule like Jesus'.

A domain general decision rule such as "Do that which maximizes your inclusive fitness" cannot guide behavior in ways that actually do maximize fitness, because what counts as fit behavior differs from domain to domain. Therefore, like the Talmud, psychological mechanisms governing evolutionarily important domains of human activity must be domain specific.

The easiest way to see that Darwinian algorithms must be domain specific is to ask whether the opposite is possible: In theory, could one construct a domain general, content-independent decision rule, that, for any two courses of action, would evaluate which better serves the end of maximizing inclusive fitness?

Such a rule must include a criterion for assessing inclusive fitness: there must be some observable environmental variable against which courses of action from any domain of human activity can be measured. As the maximization of inclusive fitness means differential representation of genes in subsequent generations, the time at which the consequence of an action can be assessed is remote from the time the action is taken. For simplicity's sake, let us assume that number of grandoffspring produced by the end of one's life is an adequate assessment of inclusive fitness. Using this criterion, the decision rule can be rephrased more precisely as, "Choose the course of action that will result in more grandoffspring produced by the end of one's life."

But how could one possibly evaluate alternative actions using this criterion? Consider a simple, but graphic example: Should one eat feces or fruit?

Clearly, no individual has two parallel lives to lead for purposes of comparison, identical except that he or she eats feces in one life and fruit in the other. Will trial and error work?⁶ The individual who eats feces is far more likely to contract parasites or infectious diseases, thereby incurring a large fitness cost. And if this individual instead eats fruit and leaves a certain number of grandoffspring, he or she still does not know whether eating feces would have been better: for all that individual knows, feces could be a rich food source that would greatly increase fecundity.

Does learning from others constitute a solution to the problem? Imitation is useless unless those imitated have themselves solved the problem of the adaptive regulation of behavior. If the blind lead the blind, there is no advantage in imitation. However, if others are monitored not as role models for imitation but instead as natural experiments, such monitoring does allow the comparison of alternative courses of action. However, each individual life is subject to innumerable uncontrolled and random influences that the observer would have to keep track of to make valid inferences. If the observer watches some people eat fruit, and others eat feces, and waits to see which have a larger number of grandoffspring, how would the

observer know whether these individuals' differential fitness was caused by their diet or by one of the many other things they experienced in the course of their lives? Of course, perhaps the major problem is that of time delay between action and the cue used to evaluate the action: grandoffspring produced. It is fundamentally impractical to have to wait two generations to determine the value of choices that must be made today.

Moreover, why would *others* choose to learn through trial and error while the observer does not? The population of self-experimenters would be selecting themselves out, compared to the observers who parasitize their risky experiments.

Can the use of perceptual cues solve the problem? The individual could decide to eat what smells good and avoid what smells bad. However, this method violates the assumption that the information processing system is domain general, and side-steps the "grandoffspring produced" criterion entirely. Nothing smells intrinsically bad or good; the smell of feces is attractive to dung flies. Moreover, what establishes the knowledge that foul-smelling entities should not be eaten? Admitting smell or taste preferences is admitting domain specific innate knowledge. Admitting the inference that foul-smelling or foul-tasting entities should not be ingested is admitting a domain specific innate inference.

Without domain specific knowledge such as this, what kind of mechanism could result in learning to avoid feces and ingest fruit? Even if it were possible, an individual with appropriate domain specific knowledge would enjoy a selective advantage over one who relied on "trial and possibly fatal error" (Shepard, this volume). The tendency to rely on trial and error in this domain would be selected out; domain specific Darwinian algorithms governing food choice would be selected for, and become a species-typical trait.

There is also the problem of deciding which courses of action to evaluate. The possibilities for action are infinite, and the best a truly domain general mechanism could do is generate random possibilities to be run through the inclusive fitness decision rule. When a tiger bounds toward you, what should your response be? Should you file your toenails? Do a cartwheel? Sing a song? Is this the moment to run an uncountable number of randomly generated response possibilities through the decision rule? And again, how could you compute which possibility would result in more grandchildren? The alternative: Darwinian algorithms specialized for predator avoidance, that err on the side of false positives in predator detection, and, upon detecting a potential predator, constrain your responses to flight, fight, or hiding.

The domain general "grandchildren produced" criterion fails even in these simple situations. How, then, could it work in more complicated learning situations—for example, when an action that increases your in-

clusive fitness in one domain decreases it in another? Suppose the hypothetical domain general learning mechanism somehow reached the inference that sexual intercourse is a necessary condition for producing offspring. Should the individual, then, have sex at every opportunity?

According to evolutionary theory, no. There are large fitness costs associated with, for example, incest (e.g., Shepher, 1983). Given a potential partner with a physique, personality, or resources that would normally elicit sexual desire, the information that the potential partner is close kin must inhibit sexual impulses.

How could this be learned? Again, if a female engages in incest, then loses her baby after a few months, how would she know what caused the miscarriage? Each life is a series of many events (perhaps including sex near the time of conception with nonkin as well as kin), any one of which is a potential cause. Why conclude that sex with one individual, who physically and psychologically resembles other members of his sex in many respects, caused the loss of the baby?

The need to avoid incest implies the ability spontaneously and automatically to acquire the category "kin versus nonkin" by merely observing the world—even if it were possible to learn it by engaging in incest, the fitness costs would be too high. But the "number of grandoffspring produced" decision rule cannot be used to acquire evolutionarily crucial categories through mere observation: unless a categorization scheme is used to guide behavior, it has no consequences on fitness.

Kin recognition requires Darwinian algorithms tuned to environmental cues that are correlated with kin but not with nonkin. These cues must be used in a particular way: either they must be used to match self to other, as in facial or olfactory phenotype matching, or they must categorize others directly, as when one imprints during a critical period on those with whom one was raised. There are an infinite number of dimensions that could be used to carve the environment into categories; there is no assurance that a general-purpose information processing system would ever isolate those useful for creating the kin/nonkin categorization scheme, and the "grandchildren produced" criterion cannot guide such a system toward the appropriate dimensions.

Additionally, there is the problem of generalization. Suppose the psyche somehow had correctly inferred that avoiding sex with kin had positive fitness consequences. How could one generalize this knowledge about the kin/nonkin categorization scheme to other domains of human activity? Would one, for example, avoid any interaction with kin? This would be a mistake; selectively avoiding sex with kin has positive fitness consequences, but selectively avoiding helping kin has negative fitness consequences (given a certain envelope of circumstances—Hamilton, 1964).

Thus, not only must the acquisition of the kin/nonkin categorization

scheme be guided by domain specific Darwinian algorithms, but its adaptive use for guiding behavior is also domain specific. In the sexual domain, kin must be avoided; in the helping domain, they must be helped; when one needs help, kin should be among the first to be asked (Hamilton, 1964); when one is contagiously ill, kin should be selectively avoided (Tooby, 1982). The procedural knowledge governing how one behaves toward kin must differ markedly from domain to domain. Only Darwinian algorithms with procedural knowledge specific to each of these domains can assure that one responds to kin in evolutionarily appropriate ways. Simply put, *there is no domain general criterion of fitness that could guide an equipotential learning process toward the correct set of fit responses.*

Trial-and-error learning is inadequate, not only because it is slow and unreliable, but because there is no domain-independent variable for signaling error. In the sexual domain, error = sex with kin. In the helping domain, error = not helping kin given the appropriate envelope of conditions. In the disease domain, error = infecting kin.

Consequently, there are only two ways the human mind can be built. Either

1. All innate psychological mechanisms are domain general, and therefore do not track fitness at all,

or

2. Some innate psychological mechanisms are domain specific Darwinian algorithms with procedural knowledge specialized for tracking fitness in the concrete situations hominids would have encountered as Pleistocene hunter-gatherers.

Clearly, the first alternative is no alternative at all. Unguided plasticity is evolutionarily fatal: there are an infinite number of unfit courses of action, and only a narrow envelope of fit behaviors. A psyche without Darwinian algorithms is incapable of keeping the organism within this narrow envelope. The idea that humans evolved from cognitively constrained ancestors into general problem solvers, now nearly devoid of adaptive specializations but equipped instead with generalized learning mechanisms, cannot be sustained. No one has yet been able to specify a general learning mechanism or general cognitive problem solver that has the power to solve the complex array of adaptive problems faced by humans, either in principle or in practice. Moreover, not only are more general sets of decision procedures less likely to provide correct guidance, but also they tend to be slower than sets of procedures designed to take advantage of the recurrent features of defined adaptive problems. In sum, advocates of the idea that the human mind is comprised predominantly of a set of domain general learning procedures has to explain how genes that code for such a maladap-

tive system could outcompete genes that code for existing successful adaptive specializations.

10 Darwinian Algorithms Solve the "Frame Problem"

Darwinian algorithms can be seen as schema- or frame-builders: as learning mechanisms that structure experience along adaptive dimensions in a given domain. Positing them solves the "frame problem"—which is the name artificial intelligence researchers gave to the family of problems with domain general mechanisms that emerged in their own work, and that parallel those raised in the discussion above.

Researchers in artificial intelligence have found that trial and error is a good procedure for learning only when a system already has a well-specified model of what is likely to be true of a domain, a model that includes a definition of what counts as error. Programmers call this finding the "frame problem" (e.g., Boden, 1977; Fodor, 1983). To move an object, make the simplest induction, or solve a straightforward problem, the computer must already have a sophisticated model of the domain in question: what counts as an object or stimulus, what counts as a cause, how classes of entities and properties are related, how various actions change the situation, what goal is to be achieved. Unless the learning domain is severely circumscribed and the procedures highly specialized and content-dependent—unless the programmer has given the computer what amounts to vast quantities of "innate knowledge"—the computer can move nothing, learn nothing, solve nothing. The frame problem is a concrete, empirical demonstration of the philosophical objections to the *tabula rasa*. It is also a cautionary tale for advocates of domain general, content-independent learning mechanisms.

Unfortunately, the lesson has been lost on many. Although most cognitive psychologists realize that their theories must posit some innate cognitive architecture, a quick perusal of textbooks in the field will show that these still tend to be restricted to content-independent operating system characteristics: short-term stores, domain general retrieval and storage processes, imagery buffers. Researchers who do insist on the necessity of positing content-dependent schemas or frames (e.g., Minsky, 1977; Schank and Abelson, 1977) seldom ask how these frames are built. Their approach implicitly presumes that frames are the product of experience structured only by domain general learning mechanisms⁷—yet the building of frames must also be subject to the frame problem. Even Fodor (1983), a prominent exponent of the view that the mind's innate architecture includes specialized, content-dependent modules, restricts these to what he calls "input systems": perceptual or quasi-perceptual domains like vision, hearing, and language. He doubts the existence of modules governing "central" pro-

cesses like reasoning and problem solving. Yet one wonders: Without domain specific inference processes, how can all these perceptual data be expected to guide our behavior in adaptive directions?

Restricting the mind's innate architecture to perceptual systems, a content-independent operating system, a domain general concept learning mechanism, a content-independent hypothesis testing procedure, and a small ragbag of dimensions for construing similarity might be sufficient if it did not matter what a person learned—if, for example, learning that E is the most frequently used letter in the English language were as critical to one's inclusive fitness as learning that a hungry tiger can leave a sizable hole in one's life plan. But what a person learns does matter; and not only what, but when, how reliably, and how quickly. Even more important is what a person *does* with that knowledge. The purpose of learning is, presumably, to guide behavior. Should one eat gravel? Should one engage in incest? How willing should a person be to give up the last remaining food available for feeding one's own children? Natural selection theory provides definite answers to questions like these, because the wrong decision can be shown to result in large fitness costs. How can an equipotential learning system that simply looks for relations in the world provide information about the relative value, in inclusive fitness terms, of alternative courses of action? It cannot; it has no standard for assessing it.

Cognitive psychologists can persist in advocating such systems only because they are not asking what problems the mind was designed, by natural selection, to solve. The Darwinian view is that humans have innately specified cognitive programs that allow them to pursue goals that are (or once were) correlated with their inclusive fitness. These innately specified programs cannot all be domain general. Behavior is a transaction between organism and environment; to be adaptive, specific behaviors must be elicited by evolutionarily appropriate environmental cues. Only specialized, domain specific Darwinian algorithms can ensure that this will happen.⁸

11 *The Frame Problem and So-Called "Constraints" on Learning*

Biologists and psychologists have an unfortunate tendency to refer to the properties of domain specific (but not domain general) mechanisms as "constraints." For example, the one-trial learning mechanism, discovered by Garcia and Koelling (1966), that permits a ^{rat} blue jay to associate a food taste with ^{nausea} vomiting several hours later is frequently referred to as a "biological constraint on learning." Books reporting the existence of domain specific learning mechanisms frequently have titles like *Biological Boundaries of Learning* (Seligman and Hager, 1972) or *The Tangled Wing: Biological Constraints on the Human Spirit* (Konner, 1982). This terminology is seriously misleading, because it incorrectly implies that "unconstrained" learning

mechanisms are a theoretical possibility; it implicitly denies the existence of the frame problem.

All constraints are properties, but not all properties are constraints. Calling a property a "constraint" implies that the organism would have a wider range of abilities if the constraint were to be removed.

Are a bird's wings a "constraint on locomotion"? Birds can locomote by flying or hopping. Wings are a property of birds that enables them to locomote by flying, but wings are not a "constraint on locomotion." On the contrary. Wings expand the bird's capacity to locomote—without wings, the bird can fly *and* hop. Removing a bird's wings reduces its capacity to locomote—without wings, it can hop, but not fly. Wings cannot be a constraint, because removing them does not give the bird a wider range of locomoting abilities. If anything, wings should be called "enablers," because they enable an additional form of locomotion. Having them expands the bird's capacity to locomote.

A thick rubber band placed in such a way that it pins a bird's wings to its body is a constraint on the bird's ability to locomote: With the rubber band the bird can only hop; without it the bird can both fly and hop.

Similarly, there is no evidence that the domain specific mechanisms that permit one-trial learning of an association between taste and ~~vomiting~~ ^{nausea} are "constraints on learning." Removing the specific properties that allow the efficient learning of this particular association would not expand the bird's ^{rat's} capacity to learn; it would reduce it. Not only would the ~~blue jay~~ ^{blue jay} be unable to associate a food taste with an electric shock; it would also be unable to associate a food taste with ~~vomiting~~ ^{nausea}.

The tendency to refer to such innate knowledge as "constraints on learning" is perhaps the result of the mistaken notion that a *tabula rasa* is possible, that learning is possible in the absence of a great deal of domain specific innate knowledge. If true, then a property that "prepares" an organism to associate a taste with vomiting might preclude it from associating a taste with an electric shock. However, if an organism with this prepared association also had a domain general associative mechanism, there is no a priori reason why that mechanism should not work to pair taste with electric shocks. In order to call the prepared association a "constraint" on the learning caused by the general purpose mechanism, one would have to demonstrate empirically that the activation of the prepared association by the presence of food somehow causes the general-purpose mechanism to shut down.

Rozin and Schull (1987) have pointed out another way in which the terminology of constraint is misleading: it implies that the human mind was "built down" from a more general-purpose cognitive system present in our ancestors. Yet such a phylogenetic history seems far from likely: it pre-

sumes that our primate ancestors had a capacity to learn that was broader and more powerful than our own.

12 Conclusions

Many evolutionary biologists seem to think that once they have identified an adaptive function, their job is done: specifying how the organism accomplishes the function is a trivial matter. This is comparable to thinking that once Einstein had shown that $E = mc^2$, designing a nuclear power plant was a trivial matter. Understanding what properties a cognitive program must have if it is to accomplish an adaptive function is far from trivial—it is one of the most difficult and challenging problems of this century.

There is emerging a new method, here called evolutionary psychology, which is made possible by the simultaneous maturation of evolutionary biology, paleoanthropology, and cognitive psychology. Together, these disciplines allow the discovery and principled investigation of the human psyche's innate cognitive programs. We propose that they be combined according to the following guidelines:

1. Use evolutionary theory as a starting point to develop models of adaptive problems that the human psyche had to solve.
2. Attempt to determine how these adaptive problems would have manifested themselves in Pleistocene conditions, insofar as this is possible. Recurrent environmental features relevant to the adaptive problem, including constraints and relationships that existed in the social, ecological, genetic, and physical situation of early hominids, should be specified; these constitute the conditions in which the adaptive problem arose, and further define the nature of the adaptive problem. Such features and relationships constitute the only environmental information available to whatever cognitive program evolved to solve the adaptive problem. The structure of the cognitive program must be such that it can guide behavior along adaptive paths given only the information available to it in these Pleistocene conditions.
3. Integrate the model of the adaptive problem with available knowledge of the relevant Pleistocene conditions, drawing whatever valid and useful implications can be derived from this set of constraints. Catalog the specific information processing problems that must be solved if the adaptive function is to be accomplished.

This constitutes a computational theory of the adaptive information processing problem. The computational theory is then used as a heuristic for generating testable hypotheses about the structure of the cognitive programs that solve the adaptive problem in question.

4. Use the computational theory to (a) determine whether there are design features that *any* cognitive program capable of solving the adaptive problem must have and (b) develop candidate models of the structure of the cognitive programs that humans might have evolved to solve the adaptive problem. Be sure the model proposed is, in principle, powerful enough to realize the computational theory.
5. Eliminate alternative candidate models with experiments and field observation. Cognitive psychologists have already developed an impressive array of concepts and experimental methods for tracking complex information processing systems—these should be used to full advantage. The end result is a validated model of the cognitive programs in question, together with a model of what environmental information, and other factors, these programs take as input.
6. Finally, compare the model against the patterns of manifest behavior that are produced by modern conditions. Informational inputs from modern environments should produce the patterns of manifest behavior predicted by the model of the cognitive programs already developed.

As previously discussed, some who adopt the evolutionary perspective attempt to leap directly from step one to step six, neglecting the intermediate steps, searching only for correspondences between evolutionary theory and modern manifest behavior.

Attempts to finesse a precise characterization of the cognitive programs that cause human behavior have led to a series of roadblocks in the application of evolutionary biology to the behavioral sciences. Because they leave the causal chain by which evolution influenced behavior vague and unspecified, such attempts have sown the widespread confusion that hypotheses about economics, culture, consciousness, learning, rationality, social forces, etc., constitute distinct alternative hypotheses to evolutionary or “biological” explanations. Instead, such hypotheses are more properly viewed as proposals about the structure of evolved cognitive programs and the kinds of information they take as input.

Cognitive psychology and evolutionary biology are sister disciplines. The goal of evolutionary theory is to define the adaptive problems that organisms must be able to solve. The goal of cognitive psychology is to discover the information processing mechanisms that have evolved to solve them. Alone, each is incomplete for the understanding of human behavior. Together, applied as a unified research program, they offer the promise that by using their organizing principles, the level of analysis appropriate for describing and investigating human behavior has, at last, been found.

Notes

1. See, for example, Block (1980) or Fodor (1981) for more discussion of the nature of cognitive explanations.
2. At least of the same sex and age.
3. Of course there can be individual variation in cognitive programs, just as there is individual variation in the size and shape of stomachs: this can be true of any structure or process in a sexually recombining species, and such genetic variation constitutes the basis for "inherited" or "constitutional" psychological differences. However, because even simple cognitive programs or "mental organs" must contain a large number of processing steps, and so must have complex polygenic bases, they necessarily evolve slowly, leading to variation being mostly "superficial." There is a large amount of variation among humans concerning single or quantitative characteristics of specific organ systems, but there is almost *no variation* among humans in what organs exist, or the basic design of each organ system. Everyone has a heart, and a liver, and so on, and everyone's heart and liver function in much the same way. We expect that this pattern holds for "mental organs" as well. Such variation, whether it is of "physical" or "mental" organ systems, can modify the functioning of these systems between individuals—sometimes drastically. Phenylketonuria is the result of a single gene modification. Nevertheless, such variation must be recognized as modifications of a design whose integrity is largely intact, and is not likely to consist of a wholly different design, differing "from the ground up." We find implausible, on the basis of population genetics considerations, the notion that different humans have fundamentally different and competing cognitive programs, resting on wholly different genetic bases.

For these and other reasons, we believe such variation can be better detected and understood if behavioral scientists devote most of their early research effort to elucidating the most commonly shared and basic design features of human cognitive programs.

A more likely kind of phenomenon is one in which wholly different cognitive programs become *activated* in different individuals, but exist latently in all individuals, based on a species-typical genetic basis. Such facultative programs can be differentially activated early in the life cycle (setting individuals along different developmental tracks), by short-term situational elicitation, or even as the result of superficial (in the sense discussed above) genetic differences in other parts of the genome (e.g., constitutional differences or gender). Gender is the most dramatic example of this facultative latency: although the profound differences between male and female have a large genetic basis, each gender has the full genetic specification for both genders. Which set of simultaneously coexisting genes becomes activated in any particular individual depends on the presence or absence of a single gene, the H-Y antigen, on the Y chromosome.

4. Herrnstein (1977) points out that Skinnerian learning theorists were able to avoid discussion of the cognitive mechanisms governing generalization and discrimination only by ignoring the problem. Available in the environment are an *infinite number* of dimensions that could be used for generalization and discrimination—but which does the organism actually use?
5. They have been called "adaptive specializations" by Rozin (1976), "modules" by Fodor (1983), and "cognitive competences" by Chomsky (1975). In our view, such mechanisms have two defining characteristics: (1) they are most usefully described on the cognitive level of proximate causation and (2) they are adaptations. We prefer "Darwinian algorithms" to the other terms because it emphasizes both characteristics.
6. The argument holds whether you characterize the process as trial and error, induction, or hypothesis testing.
7. Recently, this belief was stated explicitly by Cheng and Holyoak (1985), who cite

"induction" as the process that builds their content-dependent "pragmatic reasoning schemas."

8. We would like to direct the reader to Rozin (1976), Herrnstein (1977), Staddon (1987), and Symons (1987) for similar arguments from slightly different perspectives.

Bibliography

- Atherton, M., and R. Schwartz (1974). Linguistic Innateness and Its Evidence. *The Journal of Philosophy* LXXI: 6.
- Block, N. (1980). What Is Functionalism? In N. Block (ed.), *Readings in Philosophy of Psychology*, Cambridge, MA: Harvard University Press.
- Boden, M. (1977). *Artificial Intelligence and Natural Man*. New York: Basic Books.
- Bruner, J. S. (1973). *Beyond the Information Given* (J. M. Anglin, ed.). New York: Norton.
- Cheng, P. W. and K. J. Holyoak (1985). Pragmatic Reasoning Schemas. *Cognitive Psychology* 17:391-416.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N. (1959). Review of Skinner's "Verbal Behavior." *Language* 35:26-58.
- Chomsky, N. (1975). *Reflections on Language*. New York: Random House.
- Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.
- Cosmides, L. (1985). Deduction or Darwinian Algorithms?: An Explanation of the "Elusive" Content Effect on the Wason Selection Task. Doctoral dissertation, Department of Psychology and Social Relations, Harvard University.
- Fodor, J. A. (1981). The Mind-Body Problem. *Scientific American* 244(1):124-133.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Garcia, J., and R. A. Koelling (1966). Relations of Cue to Consequence in Avoidance Learning. *Psychonomic Science* 4:123-124.
- Hamilton, W. D. (1964). The Genetical Evolution of Social Behavior. *Journal of Theoretical Biology* 7:1-52.
- Herrnstein, R. J. (1977) The Evolution of Behaviorism. *American Psychologist* 32:593-603.
- Hume, D. (1977/1748). *An Enquiry Concerning Human Understanding* (E. Steinberg, ed.). Indianapolis: Hackett.
- Kant, I. (1966/1781). *Critique of Pure Reason*. New York: Anchor Books.
- Katz, J. J. (1975). Innate Ideas. In S. P. Stich (ed.), *Innate Ideas*, Berkeley: University of California Press.
- Konner, M. (1982). *The Tangled Wing: Biological Constraints on the Human Spirit*. New York: Holt, Rinehart, & Winston.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.
- Marr, D., and H. K. Nishihara (1978). Visual Information Processing: Artificial Intelligence and the Sensorium of Sight. *Technology Review* October: 28-49.
- Marshall, J. C. (1981). Cognition at the Crossroads. *Nature* 289:613-614.
- Maynard Smith, J., and G. R. Price (1973). The Logic of Animal Conflict. *Nature (London)* 246:15-18.
- Minsky, M. (1977). Frame-System Theory. In P. N. Johnson-Laird and P. C. Wason (eds.), *Thinking: Readings in Cognitive Science.*, Cambridge: Cambridge University Press.
- Popper, K. R. (1972). *Objective Knowledge: An Evolutionary Approach*. London: Oxford University Press.
- Putnam, H. (1967). The "Innateness Hypothesis" and Explanatory Models in Linguistics. *Synthese* 17:12-22.
- Quine, W. V. O. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.

- Rozin, P. (1976). The Evolution of Intelligence and Access to the Cognitive Unconscious. In J. M. Sprague and A. N. Epstein (eds.), *Progress in Psychobiology and Physiological Psychology*, New York: Academic Press.
- Rozin, P., and J. Schull (1987). The Adaptive-Evolutionary Point of View in Experimental Psychology. In R. C. Atkinson, R. J. Herrnstein, G. Lindsey, and R. D. Luce (eds.), *Handbook of Experimental Psychology*.
- Schank, R., and R. P. Abelson (1977). *Scripts, Plans, Goals, and Understanding*. Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Seligman, M. E. P., and J. L. Hager (1972). *Biological Boundaries of Learning*. New York: Meredith.
- Shepard, R. N. (1981). Psychophysical Complementarity. In M. Kubovy and J. R. Pomerantz, *Perceptual Organization*, Hillsdale, N.J.: Erlbaum.
- Shepard, R. N. (1984). Ecological Constraints on Internal Representation: Resonant Kinematics of Perceiving, Imagining, Thinking, and Dreaming. *Psychological Review* 91: 417-447.
- Shepard, R. N. (this volume). Evolution of a Mesh between Principles of the Mind and Regularities of the World. In J. Dupré (ed.), *The Latest on the Best*, Cambridge, MA: MIT Press.
- Shepher, J. (1983). *Incest: A Biosocial View*. New York: Academic Press.
- Staddon, J. E. R. (1987). Learning as Inference. In R. C. Bolles and M. D. Beecher (eds.), *Evolution and Learning*, Hillsdale, NJ: Erlbaum.
- Staddon, J. E. R. (this volume), Optimality Theory and Behavior. In J. Dupré (ed.), *The Latest on the Best*, Cambridge, MA: MIT Press.
- Symons, D. (1987). If We're All Darwinians, What's the Fuss About? In C. Crawford, D. Krebs, and M. Smith (eds.), *Sociobiology and Psychology*, Hillsdale, NJ: Erlbaum.
- Tooby, J. (1982). Pathogens, Polymorphism and the Evolution of Sex. *Journal of Theoretical Biology* 97:557-576.
- Tooby, J. (1985). The Emergence of Evolutionary Psychology. In D. Pines (ed.), *Emerging Syntheses in Science*, Santa Fe: Santa Fe Institute.
- Tooby, J., and L. Cosmides (1987). Evolutionary Psychology and the Generation of Culture. Part I: Theoretical Considerations. *Ethology and Sociobiology*.
- Tooby, J., and I. DeVore (1987). The Reconstruction of Hominid Behavioral Evolution through Strategic Modeling. In W. G. Kinzey (ed.), *Primate Models for the Origin of Human Behavior*, New York: SUNY Press.
- Trivers, R. L. (1972). Parental Investment and Sexual Selection. In B. Campbell (ed.), *Sexual Selection and the Descent of Man 1871-1971*, Chicago: Aldine.
- Trivers, R. L. (1974). Parent-Offspring Conflict. *American Zoologist* 14: 249-264.
- Wanner, E., and L. R. Gleitman (1982). *Language Acquisition: The State of the Art*. Cambridge: Cambridge University Press.
- Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. Cambridge, MA: Harvard University Press.