From: The Cognitive Neurosciences.
M. Gazzaniga, ed. 1995.
MIT Press, Cambridge, MA.

79 From Function to Structure: The Role of Evolutionary Biology and Computational Theories in Cognitive Neuroscience

LEDA COSMIDES AND JOHN TOOBY

ABSTRACT The cognitive neuroscience of central processes is currently a mystery. The brain is a vast and complex collection of functionally integrated circuits. Recognizing that natural selection engineers a fit between structure and function is the key to isolating these circuits. Neural circuits were designed to solve adaptive problems. If one can define an adaptive problem closely enough, one can see which circuits have a structural design that is capable of solving that problem. Evolutionary biologists have developed a series of sophisticated models of adaptive problems. Some of these models analyze constraints on the evolution of the cognitive processes that govern social behavior: cooperation, threat, courtship, kin-directed assistance, and so on. These forms of social behavior are generated by complex computational machinery. To discover the functional architecture of this machinery, cognitive neuroscientists will need the powerful inferential tools that evolutionary biology provides, including its well-defined theories of adaptive function.

The cognitive sciences have been conducted as if Darwin never lived. Their goal is to isolate functionally integrated subunits of the brain and determine how they work. Yet most cognitive scientists pursue that goal without any clear notion of what "function" means in biology. When a neural circuit is discovered, very few researchers ask what its adaptive function is. Even fewer use theories of adaptive function as tools for discovering heretofore unknown neural systems. Indeed, many people in our field think that theories of adaptive function are an explanatory luxury—fanciful, unfalsifiable speculations that one indulges in at the end of a project, after the hard work of figuring out the structure of a circuit has been done.

LEDA COSMIDES Department of Psychology, and JOHN TOOBY Department of Anthropology, Center for Evolutionary Psychology, University of California, Santa Barbara, Calif.

In this chapter, we will argue that theories of adaptive function are not a luxury. They are a necessity, crucial to the future development of cognitive neuroscience. Without them, cognitive neuroscientists will not know what to look for and will not know how to interpret their results. As a result, they will be unable to isolate functionally integrated subunits of the brain.

Explanation and discovery in cognitive neuroscience

> [t]rying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers: it just cannot be done. In order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and the different shapes of birds' wings make sense. (Marr, 1982, 27)

David Marr developed a general explanatory system for cognitive science that is much cited but rarely applied. His three-level system applies to any device that processes information—a calculator, a cash register, a television, a computer, a brain. It is based on the following observations:

1. Information-processing devices are designed to solve problems.
2. They solve problems by virtue of their structure.

3. Hence, to explain the structure of a device, one needs to know

a. *what* problem it was designed to solve, and

b. *why* it was designed to solve that problem and not some other one.

In other words, one must develop a task analysis of the problem, or what Marr called a *computational theory*. Knowing the physical structure of a cognitive device and the information-processing program realized by that structure is not enough. For human-made artifacts and biological systems, form follows function. The physical structure is there because it embodies a set of programs; the programs are there because they solve a particular problem. A computational theory specifies what that problem is and why there is a device to solve it. It specifies the *function* of an information-processing device. Marr felt that the computational theory was the most important and the most neglected level of explanation in the cognitive sciences.

This functional level of explanation has not been neglected in the biological sciences, however, because it is essential for understanding how natural selection designs organisms (for background, see chapter 78). An organism's phenotypic structure can be thought of as a collection of design features—of machines, such as the eye or liver. A design feature can cause its own spread by solving adaptive problems—problems, such as detecting predators or detoxifying poisons, that recur over many generations and whose solution tends to promote reproduction. Natural selection is a feedback process that "chooses" among alternative designs on the basis of how well they function. By selecting designs on the basis of how well they solve adaptive problems, this process engineers a tight fit between the function of a device and its structure. To understand this causal relationship, biologists had to develop a theoretical vocabulary that distinguishes between structure and function. Marr's computational theory is a functional level of explanation that corresponds roughly to what biologists refer to as the "ultimate" or "functional" explanation of a phenotypic structure.

Even though there is a close causal relationship between the function of an information-processing device and its structure, a computational theory of a device does not uniquely specify its structure. This is because there are many ways to skin a cat. More precisely:

4. Many different information-processing programs can solve the same problem. These programs may differ in how they represent information, in the processes whereby they transform input into output, or both. So knowing the goal of a computation does not uniquely determine the design of the program that realizes that goal in the device under consideration.

5. Many different physical systems—from neurons in a brain to silicon chips in a computer—can implement the same program.[1] So knowing the structure of a program does not uniquely determine the properties of the physical system that implements it. Moreover, the same physical system can implement many programs, so knowing the physical properties of a system cannot tell one which programs it implements (table 79.1)

A computational theory defines what problem the device solves and why it solves it, but it does not specify *how* this is accomplished; theories about programs and their physical substrate specify how the device solves the problem. Each explanatory level addresses a different question. To understand an information-processing device completely, Marr argued, one needs explanations on all three levels: computational theory, programming, and hardware (see table 79.1).

A computational theory of function is more than an explanatory luxury, however. It is an essential tool for discovery in the cognitive and neural sciences. Whether a mechanism was designed by natural selection or

TABLE 79.1

Three levels at which any machine carrying out an information-processing task must be understood

1. *Computational theory:*
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?
2. *Representation and algorithm:*
How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?
3. *Hardware implementation:*
How can the representation and algorithm be realized physically?
In evolutionary biology:
Explanations at the level of the computational theory are called *ultimate*-level explanations.
Explanations at the level of representation and algorithm, or at the level of hardware implementation, are called *proximate*-level explanations.

From Marr, 1982, 25.

by the intentional actions of a human engineer, one can count on there being a close causal relationship between its structure and its function. A theory of function may not determine a program's structure uniquely, but it reduces the number of possibilities to an empirically manageable number. Task demands radically constrain the range of possible solutions; consequently, very few cognitive programs are capable of solving any given adaptive problem. By developing a careful task analysis of an information-processing problem, one can vastly simplify the empirical search for the cognitive program that solves it. And once that program has been identified, it is easy to develop clinical tests that will target its neural basis.

It is currently fashionable to think that the findings of neuroscience will eventually place strong constraints on theory formation at the cognitive level. In this view, once we know enough about the properties of neurons, neurotransmitters, and cellular development, figuring out what cognitive programs the human mind contains will become a trivial task. This cannot be true. There are millions of animal species on earth, each with a different set of cognitive programs. *The same basic neural tissue embodies all of these programs.* Facts about the properties of neurons, neurotransmitters, and cellular development cannot tell one which of these millions of programs the human mind contains.

The cognitive structure of an information-processing device "depends more upon the computational problems that have to be solved than upon the particular hardware in which their solutions are implemented" (Marr, 1982, 27). In other words, knowing *what* and *why* allows one to generate focused hypotheses about *how*. To figure out how the mind works, cognitive neuroscientists will need to know what problems our cognitive and neural mechanisms were designed to solve.

Beyond intuition: How to build a computational theory

To illustrate the notion of a computational theory, Marr asks us to consider the *what* and *why* of a cash register at a checkout counter in a grocery store. We know the *what* of a cash register: It adds numbers. Addition is an operation that maps pairs of numbers onto single numbers, and it has certain abstract properties, such as commutativity and associativity (table 79.2). How the addition is accomplished is quite irrelevant: Any set of representations and algorithms that satisfies these abstract constraints will do. The input to the cash register is prices, which are represented by numbers. To compute a final bill, the cash register adds these numbers together. That's the *what*.

But *why* was the cash register designed to add the

TABLE 79.2
Why cash registers add

Rules defining addition	Rules governing social exchange in a supermarket
There is a unique element, "zero"; Adding zero has no effect: $2 + 0 = 2$	If you buy nothing, it should cost you nothing; and buying nothing and something should cost the same as buying just the something. (The rules of zero)
Commutativity: $(2 + 3) = (3 + 2) = 5$	The order in which goods are presented to the cashier should not affect the total. (Commutativity)
Associativity: $(2 + 3) + 4 = 2 + (3 + 4)$	Arranging the goods into two piles and paying for each pile separately should not affected the total amount you pay. (Associativity; the basic operation for combining prices)
Each number has a unique inverse that when added to the number gives zero: $2 + (-2) = 0$	If you buy an item and then return it for a refund, your total expenditure should be zero. (Inverses)

Adapted from Marr, 1982, 22–23.

prices of each item? Why not multiply them together, or subtract the price of each item from 100? According to Marr, "the reason is that the rules we *intuitively feel to be appropriate* for combining the individual prices in fact define the mathematical operation of addition" (p. 22, emphasis added). He formulates these intuitive rules as a series of constraints on how prices should be combined when people exchange money for goods, then shows that these constraints map directly onto those that define addition (see table 79.2). On this view, cash registers were designed to add because addition is the mathematical operation that realizes the constraints on buying and selling that our intuitions deem appropriate. Other mathematical operations are inappropriate because they violate these intuitions; for example, if the cash register substracted each price from 100, the more goods you chose the less you would pay—and if you chose enough goods, the store would pay *you*.

In this particular example, the buck stopped at intuition. But it shouldn't. Our intuitions are produced by the human brain, an information-processing device that was designed by the evolutionary process. To discover the structure of the brain, one needs to know *what* problems it was designed to solve and *why* it was designed to solve those problems rather than some others. In other words, one needs to ask the same questions of the brain as one would of the cash register. Cognitive science is the study of the design of minds, regardless of their origin. Cognitive neuroscience is the study of the design of minds that were produced by the evolutionary process. Evolution produced the what, and evolutionary biology is the study of why. Most cognitive neuroscientists know this. What they don't yet know is that understanding the evolutionary process can bring the architecture of the mind into sharper relief. For biological systems, the nature of the designer carries implications for the nature of the design.

The brain can process information because it contains complex neural circuits that are functionally organized. The only component of the evolutionary process that can build complex structures that are functionally organized is natural selection. And the only kind of problems that natural selection can build complexly organized structures for solving are adaptive problems (Williams, 1966; Dawkins, 1986; Tooby and Cosmides, 1990, 1992, this volume). Bearing this in mind, let us consider the source of Marr's intuitions about the cash register. Buying food at a grocery store is a form of social exchange—cooperation between two or more individuals for mutual benefit. The adaptive problems that arise when individuals engage in this form of cooperation have constituted a long-enduring selection pressure on the hominid line. Paleoanthropological evidence indicates that social exchange extends back at least two million years in the human line, and the fact that social exchange exists in some of our primate cousins suggests that it may be even more ancient than that. It is exactly the kind of problem that selection can build cognitive mechanisms for solving.

Social exchange is not a recent cultural invention, like writing, yam cultivation, or computer programming; if it were, one would expect to find evidence of its having one or several points of origin, of its having spread by contact, and of its being extremely elaborated in some cultures and absent in others. But its distribution does not fit this pattern. Social exchange is both universal and highly elaborated across human cultures, presenting itself in many forms: reciprocal gift-giving, food sharing, market pricing, and so on (Cosmides and Tooby, 1992; Fiske, 1992). It is an ancient, pervasive, and central part of human social life.

The computational mechanisms that give rise to social exchange behavior in a species must satisfy certain *evolvability constraints*. Selection cannot construct mechanisms in any species—including humans—that systematically violate these constraints. In evolutionary biology, researchers such as Robert Trivers, W. D. Hamilton, and Robert Axelrod have explored constraints on the evolution of social exchange using game theory, modeling it as a repeated Prisoner's Dilemma. These analyses have turned up a number of important features of this adaptive problem, a crucial one being that social exchange cannot evolve in a species unless individuals have some means of detecting individuals who cheat and excluding them from future interactions (e.g., Williams & Williams, 1957; Trivers, 1971; Axelrod and Hamilton, 1981; Axelrod, 1984; Boyd, 1988).

Behavioral ecologists have used these constraints on the evolution of social exchange to build computational theories of this adaptive problem—theories of what and why. These theories have provided a principled basis for generating hypotheses about the phenotypic design of mechanisms that generate social exchange in a variety of species. They spotlight design features that any cognitive program capable of solving this adaptive problem must have. By cataloging these design features, animal behavior researchers were able to look for—and discover—previously unknown aspects of the psychology of social exchange in species from chimpanzees, baboons and vervets to vampire bats and hermaphroditic coral-reef fish (e.g., Smuts, 1986; de Waal and Luttrell, 1988; Fischer, 1988; Wilkinson, 1988, 1990). This research strategy has been successful for a very simple reason: Very few cognitive programs satisfy the evolvability constraints for social exchange. If a species engages in this behavior (and not all do), then its cognitive architecture must contain one of these programs.

In our own species, social exchange is a universal, species-typical trait with a long evolutionary history. We have strong and cross-culturally reliable intuitions about how this form of cooperation should be conducted, which arise in the absence of any explicit instruction (Cosmides and Tooby, 1992; Fiske, 1992). In developing his computational theory of the cash register—a tool used in social exchange—David Marr was consulting these deep human intuitions.[3]

From these facts, we can deduce that the human cognitive architecture contains programs that satisfy the evolvability constraints for social exchange. As cognitive scientists, we should be able to specify what rules govern human behavior in this domain, and why we humans reliably develop circuits that embody these rules rather than others. In other words, we should be able to develop a computational theory of the organic information-processing device that governs social exchange in humans.

The empirical advantages of using evolutionary biology to develop computational theories of adaptive problems had already been amply demonstrated in the study of animal minds (e.g., Gould, 1982; Krebs and Davies, 1987; Gallistel, 1990; Real, 1991). We wanted to test its utility for studying the human mind. A powerful way of doing this would be to use an evolutionarily derived computational theory to discover cognitive mechanisms whose existence no one had previously suspected. By using evolvability constraints, we developed a computational theory of social exchange (Cosmides, 1985; Cosmides and Tooby, 1989). It suggested that the cognitive processes that govern human reasoning might have a number of design features specialized for reasoning about social exchange—what Gallistel (this volume; also Rozin, 1976) calls *adaptive specializations*.

The goal of our research is to recover, out of carefully designed experimental studies, high-resolution "maps" of the intricate mechanisms that collectively constitute the human mind. Our evolutionarily derived computational theory of social exchange has been allowing us to do that. It led us to predict a large number of design features in advance—features that no one was looking for and that most of our colleagues thought were outlandish. Experimental tests have confirmed the presence of all the design features that have been tested for so far. Those design features that have been tested and confirmed are listed in table 79.3, along with the alternative by-product hypotheses that we and our colleagues have eliminated. So far, no known theory invoking general-purpose cognitive processes has been able to explain the very precise and unique pattern of data that tests like these have generated. The data are best explained by the hypothesis that humans reliably develop circuits that are complexly specialized for reasoning about reciprocal social interactions. Parallel lines of investigation indicate that humans have also evolved additional, differently structured circuits that are specialized for reasoning about aggressive threats and protection from hazards (e.g., Manktelow and Over, 1990; Tooby and Cosmides, 1989). We are now planning clinical tests to find the neural basis for these mechanisms. By studying patient populations with autism and other neurological impairments of social cognition, we should be able to see whether dissociations occur along the fracture lines suggested by our various computational theories. (For a description of the relevant social exchange experiments, see Cosmides, 1985, 1989; Cosmides and Tooby, 1992; Gigerenzer and Hug, 1992.)

Since Marr, cognitive scientists have become familiar with the notion of developing computational theories to study perception and language, but the notion that one can develop computational theories to study the information-processing devices that give rise to social behavior is still quite alien. Yet some of the most important adaptive problems our ancestors had to solve involved negotiating the social world, and some of the best work in evolutionary biology is devoted to analyzing constraints on the evolution of mechanisms that solve these problems. There are many reasons for the neglect of these topics in the study of humans, but the primary one is that cognitive scientists have been relying on their intuitions for hypotheses rather than asking themselves what kind of problems the mind was designed to solve. Evolutionary biology addresses that question. Consequently, evolutionary

TABLE 79.3

*Reasoning about social exchange: Evidence of special design**

The following design features were predicted and found:	A number of by-product hypotheses were empirically eliminated. It was shown that:
The algorithms governing reasoning about social contracts operate even in unfamiliar situations.	Familiarity cannot explain the social contract effect.
The definition of cheating that these algorithms embody depends on one's perspective.	It is not the case that social contract content merely facilitates the application of the rules of inference of the propositional calculus.
They are just as good at computing the cost-benefit representation of a social contract from the perspective of one party as from the perspective of another.	Social contract content does not merely "afford" clear thinking.
They embody implicational procedures specified by the computational theory.	Permission schema theory cannot explain the social contract effect; in other words, application of a generalized deontic logic cannot explain the results.
They include inference procedures specialized for cheater detection.	It is not the case that any problem involving payoffs will elicit the detection of violations.
Their cheater-detection procedures cannot detect violations of social contracts that do not correspond to cheating.	
They do not include altruist detection procedures.	
They cannot operate so as to detect cheaters unless the rule has been assigned the cost-benefit representation of a social contract.	

*To show that an aspect of the phenotype is an adaptation to perform a particular function, one must show that it is particularly well designed for performing that function, and that it cannot be better explained as a by-product of some other adaptation or physical law.

biology places important constraints on theory formation in cognitive neuroscience, constraints from which one can build computational theories of adaptive information-processing problems.

Organism design theory

Knowing that the circuitry of the human mind was designed by the evolutionary process tells us something centrally illuminating: that, aside from those properties acquired by chance or imposed by engineering constraints, the mind consists of a set of information-processing circuits that were designed by natural selection to solve those adaptive problems that our hunter-gatherer ancestors faced generation after generation (see chapter 78). The better we understand the evolutionary process, adaptive problems, and ancestral life, the more intelligently we can explore and map the intricacies of the human mind.

Figuring out the structure of an organism is an exercise in reverse engineering; the field of evolutionary biology summarizes our knowledge of the engineering principles that govern the design of organisms. Taken together, these principles constitute an *organism design theory*. A major activity of evolutionary biologists is the exploration and definition of adaptive problems. By combining results derived from mathematical modeling, comparative studies, behavioral ecology, paleoanthropology, and other fields, evolutionary biologists try to identify what problems the mind was designed to solve and why it was designed to solve those problems rather than some other ones. In other words, they explore exactly those questions that Marr argued were essential for developing computational theories of adaptive information-processing problems.

Computational theories address what and why, but because there are multiple ways of achieving any solution, they are not sufficient to specify how. But the more closely one can define what and why—the more one can constrain what would count as a solution—the more clearly we can see which hypotheses about mechanisms are viable and which are not. The more constraints one can discover, the more the field of possible solutions is narrowed, and the more one can concentrate empirical efforts on discriminating between viable hypotheses.

Natural selection is capable of producing only certain kinds of designs: designs that have promoted their own reproduction in past environments. It constrains what counts as an adaptive problem, and therefore narrows the field of possible solutions. In evolutionary analyses, cognitive scientists will discover a rich and surprisingly powerful source of constraints from which precise computational theories can be built. Indeed, these analyses provide the only source of constraints for the cognitive processes that govern human social behavior. Table 79.4 lists families of constraints that cognitive scientists could be using, but are not.

TABLE 79.4

Evolutionary biology provides constraints from which computational theories of adaptive information-processing problems can be built

To build a computational theory, one must answer two questions:

1. What is the adaptive problem?

2. What information would have been available in ancestral environments for solving it?

Some sources of constraints

1. More precise definition of Marr's "goal" of processing that is appropriate to evolved (as opposed to artificial) information-processing systems

2. Game-theoretic models of the dynamics of natural selection (e.g., kin selection, Prisoner's Dilemma, and cooperation—particularly useful for analysis of cognitive mechanisms responsible for social behavior)

3. Evolvability constraints: Can a design with properties X, Y, and Z evolve, or would it have been selected out by alternative designs with different properties? (i.e., does the design represent an evolutionarily stable strategy?—related to point 2)

4. Hunter-gatherer studies and paleoanthropology—source of information about the environmental background against which our cognitive architecture evolved (Information that is present now may not have been present then, and vice versa.)

5. Studies of the algorithms and representations whereby other animals solve the same adaptive problem (These will sometimes be the same, sometimes different.)

We would like to illustrate this point with an extended example involving social behavior. Consider Hamilton's rule, which describes the selection pressures operating on mechanisms that generate behaviors that have a reproductive impact on an organism and its kin (Hamilton, 1964). The rule defines (in part) what counts as biologically successful outcomes in these kinds of situations. These outcomes often cannot be reached unless specific information is obtained and processed by the organism.

In the simplest case of two individuals, a mechanism that produces acts of assistance has an evolutionary advantage over alternative mechanisms if it reliably causes individual i to help relative j whenevener $C_i < r_{i,j}B_j$. In this equation, C_i is the cost to i of rendering an act of assistance to j, measured in terms of foregone reproduction; B_j is the benefit to j of receiving that act of assistance, measured in terms of enhanced reproduction; and $r_{i,j}$ is the probability that a randomly sampled gene will be present at the same locus in the relative due to joint inheritance from a common ancestor.

Other things being equal, the more closely the behaviors produced by cognitive mechanisms conform to Hamilton's rule, the more strongly those mechanisms will be selected for. A design feature that systematically caused an individual to help more than this—or less than this—would be selected against.

This means that the cognitive programs of an organism that confers benefits on kin cannot violate Hamilton's rule. Cognitive programs that systematically violate this constraint cannot be selected for. Cognitive programs that satisfy this constraint can be selected for. 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. Hamilton's rule is completely general: It is inherent in the dynamics of natural selection, true of any species on any planet at any time. One can call theoretical constraints of this kind *evolvability constraints*; they specify the class of mechanisms that can, in principle, evolve (Tooby and Cosmides, 1992; Cosmides and Tooby, 1994). The evolvability constraints for one adaptive problem usually differ from those for another.

Under many ecological conditions, this selection pressure defines an information-processing problem for whose solution organisms will be selected to evolve mechanisms. Hamilton's rule answers the three questions that Marr said a computational theory of an information-processing problem should answer: It identifies the goal of a computation, why it is relevant, and the logic of the strategy by which it can be carried out (Marr, 1982, 25; see table 79.1).

Using this description of an adaptive problem as a starting point, one can immediately begin to define the cognitive subtasks that would have to be addressed by any set of mechanisms capable of producing behavior that conforms to this rule. What information-processing mechanisms evolved to reliably identify relatives, for example? What criteria and procedures to do they embody? That is, do these mechanisms define an individual as a sibling if that individual (a) was nursed by the same female who nursed you, (b) resided in close contact with you during your first three years of life, or (c) looks or smells similar to your mother, within a certain error tolerance? What kind of information is processed to estimate $r_{i,j}$, the degree of relatedness? Under ancestral conditions, did siblings and cousins coreside, such that one might expect the evolution of mechanisms that discriminate between the two? After all, $r_{i,\text{full sib}} = 4r_{i,\text{first cousin}}$. What kind of mechanisms would have been capable of estimating the magnitudes of the consequences of specific actions on one's own and on others' reproduction? (For example, the estimation procedures of vampire bats could be tied directly to volume of regurgitated blood fed to a relative, as this is the only form of help they give.) What kinds of decision rules combine these various pieces of information to produce behavior that conforms to Hamilton's rule? And so on.

This example highlights several points about the connection between evolutionary biology and the cognitive sciences:

1. *Knowledge drawn from evolutionary biology can be used to discover previously unknown functional organization in our cognitive architecture.* Hamilton's rule is not intuitively obvious; researchers would not look for cognitive mechanisms that are well designed for producing behavior that conforms to this rule unless they had already heard of it. After Hamilton's rule had been formulated, behavioral ecologists began to discover psychological mechanisms that embodied it in many nonhuman animals (Krebs and Davies, 1987). Unguided empiricism is unlikely to uncover a mechanism that is well designed to solve a problem of this kind.

2. *By using the definition of an adaptive problem, one can easily generate hypotheses about the design features of information-processing mechanisms, even when these mechanisms are designed to produce social behavior.* Knowing the definition of the problem allows one to break it down into cognitive subtasks, such as kin recognition, kin categorization, and cost-benefit estimation, in the same way that knowing that the adaptive function of the visual system is scene analysis allows one to identify subtasks such as depth perception and color constancy.

3. *Knowing the ancestral conditions under which a species evolved can suggest fruitful hypotheses about design features of the cognitive adaptations that solve the problem.* For example, the key task in developing a computational theory of kin identification is identifying cues that would have been reliably correlated with kinship in ancestral environments without also being correlated with lack of kinship. If there are no such cues, then a kin identification mechanism cannot be selected for. If there are several possible cues, then empirical tests are the only way to determine which one(s) the system uses. Even so, considering what kind of information was available simplifies the task immensely: Coresidence is a reliable cue of sibhood in some species, but other cues would have to be picked up and processed in a species in which siblings and cousins coreside.

4. *Knowing about ancestral conditions can help one avoid conceptual wrong turns in the interpretation of data.* The cue "looks like me" is not a good candidate cue for kin identification, because our hunter-gatherer ancestors did not have mirrors. It therefore would have been difficult to form an accurate template of one's own face for comparison. If one were to find data suggesting that this cue is used, one should consider conducting tests to see whether this is an incidental correlation caused by the use of a more likely cue, such as "looks like my mother."

5. *A computational theory built from evolutionary constraints can provide a standard of good design.* A design for solving this adaptive problem can be evaluated by determining how closely it produces behavior that tracks Hamilton's rule. Standards of good design are an essential tool for cognitive scientists because they allow one to determine whether a hypothesized mechanism is capa-

ble of solving the adaptive problem in question and to decide whether that mechanism would have done a better job under ancestral conditions than alternative designs.

Some programs are not capable of solving a particular problem. Hypotheses that propose such programs should be eliminated from consideration. Cognitive scientists have developed powerful methods for determining whether a program is capable of solving a problem, but these methods can be used only if one has a detailed computational theory defining what the problem is. Two particularly powerful methods are as follows:

a. *Computational modeling*. One can implement the program on a computer, run the program, and see what happens.

b. *Solvability analysis*. Theoretical analyses can sometimes reveal that a proposed program is incapable of solving a problem. These analyses can be formal or informal. The learnability analyses used in developmental psycholinguistics are of both varieties (Pinker, 1979, 1984; Wexler and Culicover, 1980). The problem in question is how a child learns the grammar of his or her native language, given the information present in the child's environment. Mathematical or logical theorems can sometimes be used to prove that programs with certain formal properties are incapable of solving this problem. Informally, a grammar-learning program that works only if the child gets negative feedback about grammatical errors can be eliminated from consideration if one can show that the necessary feedback information is absent from the child's environment.

The use of these powerful methods has been largely restricted to the study of vision and language, where cognitive scientists have developed computational theories. But these methods can be applied to many other adaptive problems—including ones involving social behavior—if evolutionary analyses are used to develop comptuational theories of them. For example, because Hamilton's rule provides a standard of good design, it can be used to evaluate the popular assumption that "central" processes in humans are general purpose and content-free (e.g., Fodor, 1983).

Content-free systems are limited to knowing what can be validly derived by general processes from perceptual information. Imagine, then, a content-free architecture situated in an ancestral hunter-gatherer. When the individual with this architecture sees a relative, there is nothing in the stimulus array that tells her how much she should help that relative. And there is no consequence that she can observe that tells her whether, from a fitness point of view, she helped too much, not enough, or just the right amount, where "the right amount" is defined by $C_{ego} < r_{ego,j}B_j$. Ancestral environments lack the information necessary for inducing this rule ontogenetically (as do modern ones, for that matter). Even worse, the correct rule cannot be learned from others: An implication of Hamilton's rule is that selection will design circuits that motivate kin to socialize a child into behaving in ways that are contrary to the very rule that the child must induce (Trivers, 1974).

By developing a computational theory based on Hamilton's rule, one can easily see that a content-free architecture fails even an informal solvability test for this adaptive problem. And, because Hamilton's rule defines a particularly strong selection pressure, the content-free architecture also fails an evolvability test (Tooby and Cosmides, 1992; Cosmides and Tooby, 1994).

6. ***Insights from evolutionary biology can bring functional organization into clear focus at the cognitive level, but not at the neurobiological level.*** Hamilton's rule immediately suggests hypotheses about the functional organization of mechanisms described in information-processing terms, but it tells one very little about the neurobiology that implements these mechanisms—it cannot be straightforwardly related to hypotheses about brain chemistry or neuroanatomy. However, once one knows the properties of the cognitive mechanisms that solve this adaptive problem, it should be far easier to discover the structure of the neural mechanisms that implement them (see Tooby and Cosmides, 1992, and chapter 78). The key to finding functional organization at the neural level is finding functional organization at the cognitive level.

Hamilton's rule is a rich source of constraints from which to build computational theories of the adaptive problems associated with kin-directed social behavior. But it is not unique in this regard. When mathematical game theory was incorporated into evolutionary analyses, it became clear that natural selection constrains which kinds of circuits can evolve. For many domains of human activity, evolutionary biology can be used to determine what kind of circuits would have been quickly selected out, and what kind were likely to have become universal and species-typical. For this reason, knowledge of natural selection and of the ancestral environments in which it operated can be used to create computational theories of adaptive information-processing problems. Evolutionary biology provides a principled way of deciding what domains are likely to have associated modules[3] or mental organs—it allows one to pinpoint adaptive problems that the human mind must be able to solve with special efficiency, and it suggests design features that any mechanism capable of solving these problems must have. Of equal importance, evolutionary biology provides the definition of successful processing that is most relevant to the study of *biological* information processing systems: It gives technical content to the concept of function, telling the psychologist what adaptive goals our cognitive mechanisms must be able to accomplish.

The approach employed by Marr and others—developing computational theories of a problem defined in functional terms—has been very successful, especially in the field of perception, where the function or goal of successful processing is intuitively obvious. But for most kinds of adaptive problems (and, therefore, for most of our cognitive mechanisms), function is far from obvious, and intuition uninformed by modern biology is unreliable or misleading. In social cognition, for example, what constitutes adaptive or functional reasoning is a sophisticated biological problem in itself, and is not susceptible to impressionistic, ad hoc theorizing. There exists no domain-general standard for adaptation or successful processing; therefore functionality must be assessed through reference to evolutionary biology, adaptive problem by adaptive problem.

Fortunately, over the last 30 years, there have been rapid advances in the technical theory of adaptation. There are now a series of sophisticated models of what constitutes adaptive behavior in different domains of human life, especially those that involve social behavior. It is therefore possible to develop, out of particular areas of evolutionary biology, computational theories of the specialized cognitive abilities that were necessary for adaptive conduct in humans.

Conclusion

Textbooks in psychology are organized according to a folk-psychological categorization of mechanisms: attention, memory, reasoning, learning. In contrast, textbooks in evolutionary biology and behavioral ecology are organized according to adaptive problems: foraging (hunting and gathering), predator avoidance, resource competition, fighting, coalitional aggression, dominance and status, inbreeding avoidance, sexual attraction, courtship, pair-bond formation, trade-offs between mating effort and parenting effort, mating system, sexual conflict, paternity uncertainty and sexual jealousy, parental investment, discriminative parental care, reciprocal altruism, kin altruism, cooperative hunting, signaling and communication, navigation, habitat selection. Behavioral ecologists and evolutionary biologists have created a library of sophisticated models of the selection pressures, strategies, and trade-offs that characterize these adaptive problems.

Which model is applicable for a given species depends on certain key life-history parameters. Findings from paleoanthropology, hunter-gatherer archeology, and studies of the ways of life of modern hunter-gatherer populations locate humans in this theoretical landscape by filling in the critical parameter values. Ancestral hominids were savannah-living primates; omnivores, exposed to a wide variety of plant toxins and having a sexual division of labor between hunting and gathering; mammals with altricial young, long periods of biparental investment in offspring, pair-bonds, and an extended period of physiologically obligatory female investment in pregnancy and lactation. They were a long-lived, low-fecundity species in which variance in male reproductive success was higher than variance in female reproductive success. They lived in small, nomadic, kin-based bands of perhaps 50 to 100; they would rarely have seen more than 1000 people at one time; they had little opportunity to store provisions for the future; they engaged in cooperative hunting, defense, and aggressive coalitions; they made tools and engaged in extensive amounts of cooperative reciprocation; they were vulnerable to a large variety of parasites and pathogens. When these parameters are combined with formal models from evolutionary biology and behavioral ecology, a reasonably consistent picture of ancestral life begins to appear (e.g., Tooby and DeVore, 1987). In this picture, the adaptive problems posed by social life loom large. Most of these are characterized by strict evolvability constraints, which could only be satisfied by cognitive programs that are specialized for reasoning about the social world. This suggests that our evolved mental architecture contains a large and intricate "faculty" of social cognition (Brothers, 1990; Cosmides and Tooby, 1992; Fiske, 1992; Jackendoff, 1992). Yet virtually no work in cognitive neuroscience is devoted to looking for dissociations between different forms of social reasoning, or between social reasoning and other cognitive functions. The work on autism as a neurological impairment of a "theory of mind" module is a notable and very successful exception (e.g., Baron-Cohen, Leslie, and Frith, 1985; Frith, 1989; Leslie, 1987.)

Textbooks in evolutionary biology are organized according to adaptive problems because these are the only problems that selection can build mechanisms for solving. Textbooks in behavioral ecology are organized according to adaptive problems because circuits that are functionally specialized for solving these problems have been found in species after species. No less should be true of humans. To find such circuits, however, cognitive neuroscientists will need the powerful inferential tools that evolutionary biology provides.

Through the computational theory, evolutionary biology allows the matching of algorithm to adaptive problem: Evolutionary biology defines information-processing problems that the mind must be able to solve, and the task of cognitive neuroscience is to uncover the nature of the algorithms that solve them. The brain's microcircuitry was designed to implement these algorithms, so a map of their cognitive structure can be used to bring order out of chaos at the neural level.

Atheoretical approaches will not suffice—a random stroll through hypothesis space will not allow one to distinguish figure from ground in a complex system. To isolate a functionally integrated mechanism within a complex system, one needs a theory of what function that mechanism was designed to perform. Sophisticated theories of adaptive function are therefore essential if cognitive neuroscience is to flourish.

ACKNOWLEDGMENTS We thank Steve Pinker and Mike Gazzaniga for valuable discussions of the issues discussed in this chapter. We are grateful to the McDonnell Foundation for financial support, and also for NSF grant BNS9157-449 to John Tooby.

NOTES

1. For example, consider the fact that certain text editing programs, such as WordStar, have been implemented on machines with different hardware architectures. The program is the same, in the sense that functional relationships among representations are preserved. The same inputs produce the same outputs: ^G always erases a letter, ^KV always moves a block, and so on.
2. Had Marr known about the importance of cheating in evolutionary analyses of social exchange, he might have been able to understand other features of the cash register as well. Most cash registers have anticheating devices: cash drawers that lock until a new set of prices is punched in, two rolls of tape that keep track of transactions (one is for the customer; the other rolls into an inaccessible place in the cash register, preventing the clerk from altering the totals to match the amount of cash in the drawer). In a way akin to the evolutionary process, as more sophisticated technologies become available and cheap, one might expect the anticheating design features of cash registers to become more sophisticated as well.
3. We do not mean "modules" in Fodor's sense; his criteria do not lay appropriate emphasis on functional organization for solving adaptive problems.

REFERENCES

Axelrod, R., 1984. *The Evolution of Cooperation.* New York: Basic Books.

Axelrod, R., and W. D. Hamilton, 1981. The evolution of cooperation. *Science* 211:1390–1396.

Baron-Cohen, S., A. Leslie, and U. Frith, 1985. Does the autistic child have a "theory of mind"? *Cognition* 21:37–46.

Boyd, R., 1988. Is the repeated prisoner's dilemma a good model of reciprocal altruism? *Ethol. Sociobiol.* 9:211–222.

Brothers, L., 1990. The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci.* 1:27–51.

Cheney, D. L., and R. Seyfarth, 1990. *How Monkeys See the World.* Chicago: University of Chicago Press.

Cosmides, L., 1985. Deduction or Darwinian algorithms? An explanation of the "elusive" content effect on the Wason selection task. Ph.D. diss. Harvard University. University Microfilms #86-02206.

Cosmides, L., 1989. The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31:187–276.

Cosmides, L., and J. Tooby, 1989. Evolutionary psychology and the generation of culture, Part II. Case study: A computational theory of social exchange. *Ethol. Sociobiol.* 10: 51–97.

Cosmides, L., and J. Tooby, 1992. Cognitive adaptations for social exchange. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture,* J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press.

Cosmides, L., and J. Tooby, 1994. Origins of domain specificity: The evolution of functional organization. In *Mapping the Mind: Domain Specificity in Cognition and Culture,* L. Hirschfeld and S. Gelman, eds. New York: Cambridge University Press.

Dawkins, R., 1986. *The Blind Watchmaker.* New York: Norton.

De Waal, F. B. M., and L. M. Luttrell, 1988. Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethol. Sociobiol.* 9:101–118.

Fischer, E. A., 1988. Simultaneous hermaphroditism, tit-for-tat, and the evolutionary stability of social systems. *Ethol. Sociobiol.* 9:119–136.

Fiske, A. P., 1992. *Structures of Social Life: The Four Elementary Forms of Human Relations*. New York: Free Press.

Fodor, J. A., 1983. *The Modularity of Mind*. Cambridge, Mass.: MIT Press.

Frith, U., 1989. *Autism: Explaining the Enigma*. Oxford: Blackwell.

Gallistel, C. R., 1990. *The Organization of Learning*. Cambridge, Mass.: MIT Press.

Gigerenzer, G., and K. Hug, 1992. Domain-specific reasoning: Social contracts, cheating and perspective change. *Cognition* 43:127–171.

Gould, J. L., 1982. *Ethology: The Mechanisms and Evolution of Behavior*. New York: Norton.

Hamilton, W. D., 1964. The genetical evolution of social behaviour. I, II. *J. Theor. Biol.* 7:1–52.

Jackendoff, R., 1992. *Languages of the Mind*. Cambridge, Mass.: MIT Press.

Krebs, J. R., and N. B. Davies, 1987. *An Introduction to Behavioural Ecology*. Oxford: Blackwell.

Leslie, A. M., 1987. Pretense and representation: The origins of "theory of mind". *Psychol. Rev.* 94:412–426.

Manktelow, K. I., and D. E. Over, 1990. Deontic thought and the selection task. In *Lines of Thinking*, vol. 1, K. J. Gilhooly, M. T. G. Keane, R. H. Logie, and G. Erdos, eds. New York: Wiley.

Marr, D., 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.

Pinker, S., 1979. Formal models of language learning. *Cognition* 7:217–283.

Pinker, S., 1984. *Language Learnability and Language Development*. Cambridge, Mass.: Harvard University Press.

Pinker, S., 1994. *The Language Instinct*. New York: Morrow.

Real, L. A., 1991. Animal choice behavior and the evolution of cognitive architecture. *Science* 253:980–986.

Rozin, P., 1976. The evolution of intelligence and access to the cognitive unconscious. In *Progress in Psychobiology and Physiological Psychology*, J. M. Sprague and A. N. Epstein, eds. New York: Academic Press.

Smuts, B., 1986. *Sex and Friendship in Baboons*. Hawthorne, N.Y.: Aldine.

Tooby, J., and L. Cosmides, 1990. The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethol. Sociobiol.* 11:375–424.

Tooby, J., and L. Cosmides, 1992. The psychological foundations of culture. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press.

Tooby, J., and L. Cosmides, 1989. The logic of threat. Human Behavior and Evolution Society, Evanston, IL.

Tooby, J., and I. DeVore, 1987. The reconstruction of hominid behavioral evolution through strategic modeling. In *Primate Models of Hominid Behavior*, W. Kinzey, ed. New York: SUNY Press.

Trivers, R. L., 1971. The evolution of reciprocal altruism. *Q. Rev. Biol.* 46:35–57.

Trivers, R. L., 1974. Parent-offspring conflict. *Am. Zool.* 14:249–264.

Wexler, K., and P. Culicover, 1980. *Formal Principles of Language Acquisition*. Cambridge, Mass.: MIT Press.

Wilkinson, G. S., 1988. Reciprocal altruism in bats and other mammals. *Ethol. Sociobiol.* 9:85–100.

Wilkinson, G. S., 1990. Food sharing in vampire bats. *Sci. Am.* February: 76–82.

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

Williams, G. C., and D. C. Williams, 1957. Natural selection of individually harmful social adaptations among sibs with special reference to social insects. *Evolution* 11:32–39.

THE COGNITIVE NEUROSCIENCES

Michael S. Gazzaniga, *Editor-in-Chief*

Section Editors: Emilio Bizzi
Ira B. Black
Colin Blakemore
Leda Cosmides
Stephen M. Kosslyn
Joseph E. LeDoux
J. Anthony Movshon
Steven Pinker
Michael I. Posner
Pasko Rakic
Daniel L. Schacter
John Tooby
Endel Tulving

A BRADFORD BOOK
THE MIT PRESS
CAMBRIDGE, MASSACHUSETTS
LONDON, ENGLAND