Cognition, 50 (1994) 41–77
0010-0277/94/$07.00

Beyond intuition and instinct blindness: toward an evolutionarily rigorous cognitive science

Leda Cosmides*,[a], John Tooby[b]

[a]*Laboratory of Evolutionary Psychology, Department of Psychology, University of California, Santa Barbara, CA 93106-3210, USA*

[b]*Laboratory of Evolutionary Psychology, Department of Anthropology, University of California, Santa Barbara, CA 93106-3210, USA*

Abstract

Cognitive psychology has an opportunity to turn itself into a theoretically rigorous discipline in which a powerful set of theories organize observations and suggest focused new hypotheses. This cannot happen, however, as long as intuition and folk psychology continue to set our research agenda. This is because intuition systematically blinds us to the full universe of problems our minds spontaneously solve, restricting our attention instead to a minute class of unrepresentative "high-level" problems. In contrast, evolutionarily rigorous theories of adaptive function are the logical foundation on which to build cognitive theories, because the architecture of the human mind acquired its functional organization through the evolutionary process. Theories of adaptive function specify what problems our cognitive mechanisms were designed by evolution to solve, thereby supplying critical information about what their design features are likely to be. This information can free cognitive scientists from the blinders of intuition and folk psychology, allowing them to construct experiments capable of detecting complex mechanisms they otherwise would not have thought to test for. The choice is not

*Corresponding author.

For many illuminating discussions on these topics, we warmly thank Pascal Boyer, David Buss, Martin Daly, Mike Gazzaniga, Gerd Gigerenzer, Steve Pinker, Roger Shepard, Dan Sperber, Don Symons, Margo Wilson, and the members of the Laboratory of Evolutionary Psychology (UCSB). We also thank Don Symons for calling our attention to Gary Larson's "Stoppit" cartoon and, especially, Steve Pinker for his insightful comments on an earlier draft. We are grateful to the McDonnell Foundation and NSF Grant BNS9157-499 to John Tooby for their financial support.

SSDI 0010-0277(93)00589-Y

between no-nonsense empiricism and evolutionary theory: it is between folk theory and evolutionary theory.

Nothing in biology makes sense except in the light of evolution.

Theodosius Dobzhansky

Is it not reasonable to anticipate that our understanding of the human mind would be aided greatly by knowing the purpose for which it was designed?

George C. Williams

The cognitive sciences have reached a pivotal point in their development. We now have the opportunity to take our place in the far larger and more exacting scientific landscape that includes the rest of the modern biological sciences. Every day, research of immediate and direct relevance to our own is being generated in evolutionary biology, behavioral ecology, developmental biology, genetics, paleontology, population biology, and neuroscience. In turn, many of these fields are finding it necessary to use concepts and research from the cognitive sciences.

But to benefit from knowledge generated in these collateral fields, we will have to learn how to use biological facts and principles in theory formation and experimental design. This means shedding certain concepts and prejudices inherited from parochial parent traditions: the obsessive search for a cognitive architecture that is general purpose and initially content-free; the excessive reliance on results derived from artificial "intellectual" tasks; the idea that the field's scope is limited to the study of "higher" mental processes; and a long list of false dichotomies reflecting premodern biological thought – evolved/learned, evolved/developed, innate/learned, genetic/environmental, biological/social, biological/cultural, emotion/cognition, animal/human. Most importantly, cognitive scientists will have to abandon the functional agnosticism that is endemic to the field (Tooby & Cosmides, 1992).

The biological and cognitive sciences dovetail elegantly because in evolved systems – such as the human brain – there is a causal relationship between the adaptive problems a species encountered during its evolution and the design of its phenotypic structures. Indeed, a theoretical synthesis between the two fields seems inevitable, because evolutionary biologists investigate and inventory the set of adaptive information-processing problems the brain evolved to solve, and cognitive scientists investigate the design of the circuits or mechanisms that evolved to solve them. In fact, the cognitive subfields that already recognize and exploit this relationship between function and structure, such as visual perception, have made the most rapid empirical progress. These areas succeed because they are guided by (1) theories of adaptive function, (2) detailed analyses of the tasks each mechanism was designed by evolution to solve, and (3) the recognition that these tasks are usually solved by cognitive machinery that is highly functionally specialized. We believe the study of central processes can be revitalized by applying the same adaptationist program. But for this to happen, cognitive scientists will have to replace the intuitive, folk psychological notions that now dominate the field with evolutionarily rigorous theories of function.

It is exactly this reluctance to consider function that is the central impediment to the emergence of a biologically sophisticated cognitive science. Surprisingly, a few cognitive scientists have tried to ground their dismissal of functional reasoning in biology itself. The claim that natural selection is too constrained by other factors to organize organisms very functionally has indeed been made by a small number of biologists (e.g., Gould & Lewontin, 1979). However, this argument has been empirically falsified so regularly and comprehensively that it is now taken seriously only by research communities too far outside of evolutionary biology to be acquainted with its primary literature (Clutton-Brock & Harvey, 1979; Daly & Wilson, 1983; Dawkins, 1982, 1986; Krebs & Davies, 1987; Williams, 1966; Williams & Nesse, 1991).[1] Other cognitive scientists take a less ideological, more agnostic stance; most never think about function at all.

As a result, cognitive psychology has been conducted as if Darwin never lived. Most cognitive scientists proceed without any clear notion of what "function" means for biological structures like the brain, or what the explicit analysis of function could teach them. Indeed, many cognitive scientists 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 experimentation has been done.

But theories of adaptive function are not a luxury. They are an indispensable methodological tool, crucial to the future development of cognitive psychology. Atheoretical approaches will not suffice – a random stroll through hypothesis space will not allow you to distinguish figure from ground in a complex system. To isolate a functionally organized mechanism within a complex system, you need a theory of what function that mechanism was designed to perform.

This article is intended as an overview of the role we believe theories of adaptive function should play in cognitive psychology. We will briefly explain why they are important, where exactly they fit into a research program, how they bear

[1]Similar results emerge from the cognitive sciences. Although artificial intelligence researchers have been working for decades on computer vision, object recognition, color constancy, speech recognition and comprehension, and many other evolved competences of humans, naturally selected computational systems still far outperform artificial systems on the adaptive problems they evolved to solve – on those rare occasions when artificial systems can solve the assigned tasks at all. In short, natural selection is known to produce cognitive machinery of an intricate functionality as yet unmatched by the deliberate application of modern engineering. This is a far more definable standard than "optimality" – where many anti-adaptationist arguments go awry. There are an uncountable number of changes that could conceivably be introduced into the design of organisms and, consequently, the state space of potential organic designs is infinitely large and infinitely dimensioned. Thus, there is no way of defining an "optimal" point in it, much less "measuring" how closely evolution brings organisms to it. However, when definable engineering standards of functionality are applied, adaptations can be shown to be very functionally designed – for solving *adaptive* problems.

on cognitive and neural theories, and what orthodoxies they call into question. (For a more complete and detailed argument, see Tooby & Cosmides, 1992.)

I. Function determines structure

Explanation and discovery in the cognitive sciences

> . . . trying 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, p. 27)

David Marr developed a general explanatory system for the cognitive sciences 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, you need 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, you need to develop a task analysis of the problem, or what Marr called a *computational theory* (Marr, 1982). 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 is 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. An organism's phenotypic structure can be thought of as a collection of "design features" – micro-machines, such as the functional components of the eye or liver. Over evolutionary time, new design features are added or discarded from the species' design because of their consequences. A design feature will cause its own spread over generations if it has the consequence of solving adaptive problems: cross-generationally recurrent problems whose solution promotes reproduction, such as detecting predators or detoxifying poisons. 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.[2] 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.

A computational theory defines what problem the device solves and why it solves it; theories about programs and their physical substrate specify *how* the device solves the problem. "How" questions – questions about programs and hardware – currently dominate the research agenda in the cognitive sciences. Answering such questions is extremely difficult, and most cognitive scientists realize that groping in the dark is not a productive research strategy. Many see

Table 1. *Three levels at which any machine carrying out an information-processing task must be understood (from Marr, 1982, p. 25)*

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 representations and algorithm, or at the level of hardware implementation, are called *proximate* levels of explanation.

[2]All traits that comprise species-typical designs can be partitioned into adaptations, which are present because they were selected for, by-products, which are present because they are causally coupled to traits that were selected for, and noise, which was injected by the stochastic components of evolution. Like other machines, only narrowly defined aspects of organisms fit together into functional systems: most of the system is incidental to the functional properties. Unfortunately, some have misrepresented the well-supported claim that selection organizes organisms very functionally as the obviously false claim that all traits of organisms are functional – something no sensible evolutionary biologist would ever maintain. Nevertheless, cognitive scientists need to recognize that while not everything in the designs of organisms is the product of selection, all complex functional organization is (Dawkins, 1986; Pinker & Bloom, 1990; Tooby & Cosmides, 1990a, 1990b, 1992; Williams, 1966, 1985).

the need for a reliable source of theoretical guidance. The question is, what form should it take?

Why ask why? – or – how to ask how

It is currently fashionable to think that the findings of neuroscience will eventually place strong constraints on theory formation at the cognitive level. Undoubtedly they will. But extreme partisans of this position believe neural constraints will be *sufficient* for developing cognitive theories. 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. Consider the fact that there are birds that migrate by the stars, bats that echolocate, bees that compute the variance of flower patches, spiders that spin webs, humans that speak, ants that farm, lions that hunt in teams, cheetahs that hunt alone, monogamous gibbons, polyandrous seahorses, polygynous gorillas . . . 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*, and it could support many others as well. Facts about the properties of neurons, neurotransmitters, and cellular development cannot tell you which of these millions of programs the human mind contains.

Even if all neural activity is the expression of a uniform process at the cellular level, it is the arrangement of neurons – into birdsong templates or web-spinning programs – that matters. The idea that low-level neuroscience will generate a self-sufficient cognitive theory is a physicalist expression of the ethologically naive associationist/empiricist doctrine that all animal brains are essentially the same.

In fact, as David Marr put it, a program's structure "depends more upon the computational problems that have to be solved than upon the particular hardware in which their solutions are implemented" (1982, p. 27). In other words, knowing *what* and *why* places strong constraints on theories of *how*.

For this reason, a computational theory of function is not an explanatory luxury. It is an essential tool for discovery in the cognitive and neural sciences. 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, you can vastly simplify the empirical search for the cognitive program that solves it. And once that program has been identified, it becomes straightforward to develop clinical tests that will target its neural basis.

To figure out how the mind works, cognitive scientists 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 check-out 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 (see Table 2). How the addition is accomplished is quite irrelevant: any set of representations and algorithms that satisfy 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 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 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 subtracted each price from 100, the more goods you chose the less you would pay – and whenever you chose more than $100 of 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, you need to know *what* problems it was designed to solve and *why* it was designed to solve those problems rather than some other ones. In other words, you need to ask the same questions of the brain as you would of the cash register.

Cognitive science is the study of the design of minds, regardless of their origin. Cognitive psychology 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 scientists know this. What they don't yet know is that understanding the evolutionary process can bring the architecture of the mind into

Table 2. *Why cash registers add (adapted from Marr, 1982, pp. 22–23)*

Rules defining addition	Rules governing social exchange in a supermarket
1. There is a unique element, "zero"; Adding zero has no effect: 2 + 0 = 2	1. 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 for zero
2. Commutativity: (2 + 3) = (3 + 2) = 5	2. The order in which goods are presented to the cashier should not affect the total. (Commutativity.)
3. Associativity: (2 + 3) + 4 = 2 + (3 + 4)	3. Arranging the goods into two piles and paying for each pile separately should not affect the total amount you pay. (Associativity; the basic operation for combining prices.)
4. Each number has a unique inverse that when added to the number gives zero: 2 + (−2) = 0	4. If you buy an item and then return it for a refund, your total expenditure should be zero. (Inverses.)

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, where "adaptive" has a very precise, narrow technical meaning. (Dawkins, 1986; Pinker & Bloom, 1990; Tooby & Cosmides, 1990a, 1992; Williams, 1966). Bearing this in mind, let's 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 2 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, marketing-pricing, and so on (Cosmides & Tooby, 1992; Fiske, 1991). It is an ancient, pervasive and central part of human social life.

In evolutionary biology, researchers such as George Williams, 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., Axelrod, 1984; Axelrod & Hamilton, 1981; Boyd, 1988; Trivers, 1971). One can think of this as an *evolvability constraint*. Selection cannot construct mechanisms in any species – including humans – that systematically violate such constraints. Behavior is generated by computational mechanisms. If a species engages in social exchange behavior, then it does so by virtue of computational mechanisms that satisfy the evolvability constraints that characterize this adaptive problem.

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., Cheney & Seyfarth, 1990; de Waal & Luttrell, 1988; Fischer, 1988; Smuts, 1986; 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 & Tooby, 1992; Fiske, 1991). 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

[3]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 anti-cheating devices. Cash drawers lock until a new set of prices is punched in; two rolls of tape 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); and so on.

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.

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 navigating 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. In fact, these evolutionary analyses may be the *only* source of constraints available for developing computational theories of social cognition.

Principles of organic design

The field of evolutionary biology summarizes our knowledge of the engineering principles that govern the design of organisms. As a source of theoretical guidance about organic design, functionalism has an unparalleled historical track record. As Ernst Mayr notes, "The adaptationist question, 'What is the function of a given structure or organ?' has been for centuries the basis for every advance in physiology" (1983, p. 328).

Attention to function can advance the cognitive sciences as well. Aside from those properties acquired by chance or imposed by engineering constraint, the mind consists of a set of information-processing circuits that were designed by natural selection to solve adaptive problems that our hunter-gatherer ancestors faced generation after generation.[4] If we know what these problems were, we can seek mechanisms that are well engineered for solving them.

The exploration and definition of these adaptive problems is a major activity of evolutionary biologists. By combining results derived from mathematical modeling, comparative studies, behavioral ecology, paleoanthropology and other fields,

[4]Our ancestors spent the last 2 million years as Pleistocene hunter-gatherers (and several hundred million years before that as one kind of forager or another). The few thousand years since the scattered appearance of agriculture is a short stretch, in evolutionary terms (less than 1% of the past 2 million years). Complex designs – ones requiring the coordinated assembly of many novel, functionally integrated features – are built up slowly, change by change, subject to the constraint that each new design feature must solve an adaptive problem better than the previous design (the vertebrate eye is an example). For these and other reasons, it is unlikely that our species evolved complex adaptations even to agriculture, let alone to post-industrial society (for discussion, see Dawkins, 1982; Tooby & Cosmides, 1990a, 1990b).

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, evolutionary biologists explore exactly those questions that Marr argued were essential for developing computational theories of adaptive information-processing problems.

Computional theories address *what* and *why*, but because there are multiple ways of achieving any solution, experiments are needed to establish *how*. But the more precisely you can define the goal of processing – the more tightly you can constrain what would count as a solution – the more clearly you can see what a mechanism capable of producing that solution would have to look like. The more constraints you can discover, the more the field of possible solutions is narrowed, and the more you can concentrate your experimental efforts on discriminating between viable hypotheses.

A technological analogy may make this clearer. It is difficult to figure out the design of the object I'm now thinking about if all you know is that it is a machine (toaster? airplane? supercollider?). But the answer becomes progressively clearer as I add functional constraints: (1) it is well designed for entertainment (movie projector, TV, CD player?); it was not designed to project images (nothing with a screen); it is well designed for playing taped music (stereo or Walkman); it was designed to be easily portable during exercise (Walkman).

Knowing the object is well engineered for solving these problems provides powerful clues about its functional design features that can guide research. Never having seen one, you would know that it must contain a device that converts magnetic patterns into sound waves; a place to insert the tape; an outer shell no smaller than a tape, but no larger than necessary to perform the transduction; and so on.

Guessing at random would have taken forever. Information about features that have no impact on the machine's function would not have helped much either (e.g., its color, the number of scratches). Because functionally neutral features are free to vary, information about them does little to narrow your search.

Functional information helps because it narrowly specifies the outcome to be produced. The smaller the class of entities capable of producing that outcome, the more useful functional information is. This means (1) narrow definitions of outcomes are more useful than broad ones (tape player versus entertainment device), and (2) functional information is most useful when there are only a few ways of producing an outcome (Walkman versus paperweight; seeing versus scratching).

Narrow definitions of function are a powerful methodological tool for discovering the design features of any complex problem-solving device, including the human mind. Yet the definition of function that guides most research on the mind (it "processes information") is so broad that it applies even to a Walkman.

It is possible to create detailed theories of adaptive function. This is because natural selection is only capable of producing certain kinds of designs: designs that promoted their own reproduction in past environments. This rule of organic design sounds too general to be of any help. But when it is applied to real species in actual environments, this deceptively simple constraint radically limits what counts as an adaptive problem and, therefore, narrows the field of possible solutions. Table 3 lists some principles of organic design that cognitive psychologists could be using, but aren't.

Doing experiments is like playing "20 questions" with nature, and evolutionary biology gives you an advantage in this game: it tells you what questions are most worth asking, and what the answer will probably look like. It provides constraints – functional and otherwise – from which computational theories of adaptive information-processing problems can be built.

Taking function seriously

We know the cognitive science that intuition has wrought. It is more difficult, however, to know how our intuitions might have blinded us. What cognitive systems, if any, are we *not* seeing? How would evolutionary functionalism transform the science of mind?

Table 3. *Evolutionary biology provides constraints from which computational theories of adaptive information-processing problems can be built*

To build a computational theory, you need to 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)

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), kinship, predator defense, resource competition, cooperation, aggression, parental care, dominance and status, inbreeding avoidance, courtship, mateship maintenance, trade-offs between mating effort and parenting effort, mating system, sexual conflict, paternity uncertainty and sexual jealousy, signaling and communication, navigation, habitat selection, and so on.

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 prove true of humans. Twenty-first-century textbooks on human cognition will probably be organized similarly.

Fortunately, behavioral ecologists and evolutionary biologists have already created a library of sophisticated models of the selection pressures, strategies and trade-offs that characterize these very fundamental adaptive problems, which they use in studying processes of attention, memory, reasoning and learning in non-humans. Which model is applicable for a given species depends on certain key life-history parameters. Findings from paleoanthropology, hunter-gatherer archaeology, and studies of living hunter-gatherer populations locate humans in this theoretical landscape by filling in the critical parameter values. Ancestral hominids were ground-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, enduring male–female mateships, 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 20–100; they would rarely (if ever) 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 & 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 & Tooby, 1992; Fiske, 1991; Jackendoff, 1992). Yet despite its importance, very little work in the cognitive sciences has been devoted to looking for cognitive mechanisms that are specialized for reasoning about the social world. Nor have cognitive neuroscientists been looking for dissociations among 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 very notable – and very successful – exception (e.g., Baron-Cohen, Leslie & Frith, 1985; Frith, 1989; Leslie, 1987).

There are many reasons for the neglect of these topics in the study of humans (see Tooby & Cosmides, 1992), but a 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 by evolution to solve. By using evolutionary biology to remind ourselves of the types of problems hominids faced across hundreds of thousands of generations, we can escape the narrow conceptual cage imposed on us by our intuitions and folk psychology. This is not a minor point: if you don't think a thing exists, you won't take the steps necessary to find it. By having the preliminary map that an evolutionary perspective provides, we can find our way out into the vast, barely explored areas of the human cognitive architecture.

II. Computational theories derived from evolutionary biology suggest that the mind is riddled with functionally specialized circuits

During most of this century, research in psychology and the other biobehavioral and social sciences has been dominated by the assumptions of what we have elsewhere called the Standard Social Science Model (SSSM) (Tooby & Cosmides, 1992). This model's fundamental premise is that the evolved architecture of the human mind is comprised mainly of cognitive processes that are content-free, few in number and general purpose. These general-purpose mechanisms fly under names such as "learning", "induction", "imitation", "reasoning" and "the capacity for culture", and are thought to explain nearly every human phenomenon. Their structure is rarely specified by more than a wave of the hand.

In this view, the same mechanisms are thought to govern how one acquires a language and a gender identity, an aversion to incest and an appreciation for vistas, a desire for friends and a fear of spiders – indeed, nearly every thought and feeling of which humans are capable. By definition, these empiricist mechanisms have no inherent content built into their procedures, they are not designed to construct certain mental contents more readily than others, and they have no features specialized for processing particular kinds of content over others. In other words, they are assumed to operate uniformly, no matter what content, subject matter or domain of life experience they are operating on. (For this reason, such procedures are described as *content-independent*, *domain-general* or *content-free*). The premise that these mechanisms have no content to impart is what leads to the doctrine central to the modern behavioral and social sciences: that all of our particular mental content originated in the social and physical world and entered through perception. As Aquinas put this empiricist tenet a millennium ago, "There is nothing in the intellect that was not first in the senses."

As we will discuss, this view of central processes is difficult to reconcile with modern evolutionary biology.

The weakness of content-independent architectures

To some it may seem as if an evolutionary perspective supports the case that our cognitive architecture consists primarily of powerful, general-purpose problem-solvers – inference engines that embody the content-free normative theories of mathematics and logic. After all, wouldn't an organism be better equipped and better adapted if it could solve a more general class of problems over a narrower class?

This empiricist view is difficult to reconcile with evolutionary principles for a simple reason: content-free, general-purpose problem-solving mechanisms are extraordinarily weak – or even inert – compared to specialized ones. Every computational system – living or artificial – must somehow solve the frame problem (e.g., Pylyshyn, 1987). Most artificial intelligence programs have domain-specific knowledge and procedures that do this (even those that are called "general purpose"). A program equipped solely with domain-general procedures can do nothing unless the human programmer solves the frame problem for it: either by artificially constraining the problem space or by supplying the program – by fiat – with pre-existing knowledge bases ("innate" knowledge) that it could not have acquired on its own, with or without connections to a perceptual system.

However, to be a viable hypothesis about our cognitive architecture, a proposed design must pass a solvability test. It must, in principle, be able to solve problems humans are known to be able to solve. At a minimum, any proposed cognitive architecture had to produce sufficiently self-reproductive behavior in ancestral environments – we know this because all living species have been able to reproduce themselves in an unbroken chain up to the present. While artificial intelligence programs struggle to recognize and manipulate coke cans, naturally intelligent programs situated in organisms successfully negotiate through lifetimes full of biotic antagonists – predators, conspecific competitors, self-defending food items, parasites, even siblings. At the same time, these naturally intelligent programs solve a large series of intricate problems in the project of assembling a sufficient number of replacement individuals: offspring. Just as a hypothesized set of cognitive mechanisms underlying language must be able to account for the facts of human linguistic behavior, so too must any hypothetical domain-general cognitive architecture reliably generate solutions to all of the problems that were necessary for survival and reproduction in the Pleistocene. For humans and most other species, this is a remarkably diverse, highly structured and very complex set of problems.

If it can be shown that there are essential adaptive problems that humans must have been able to solve in order to have propagated and that domain-general mechanisms cannot solve them, then the domain-general hypothesis fails. We think there is a very large number of such problems, including inclusive fitness regulation, mate choice, nutritional regulation, foraging, navigation, incest avoidance, sexual jealousy, predator avoidance, social exchange – at a minimum, any kind of information-processing problem that involves motivation, and many others as well. We have developed this argument in detail elsewhere (Cosmides & Tooby, 1987, 1994; Tooby & Cosmides, 1990a, 1992), so we won't belabor it here. Instead, we will simply summarize a few of the relevant points.

(1) *The "Stoppit" problem.* There is a Gary Larson cartoon about an "all-purpose" product called "Stoppit". When sprayed from an aerosol can, Stoppit stops faucet drips, taxis, cigarette smoking, crying babies and charging elephants. An "all-purpose" cognitive program is no more feasible for an analogous reason: *what counts as adaptive behavior differs markedly from domain to domain.* An architecture equipped only with content-independent mechanisms must succeed at survival and reproduction by applying the same procedures to every adaptive problem. But there is no domain-general criterion of success or failure that correlates with fitness (e.g., what counts as a "good" mate has little in common with a "good" lunch or a "good" brother). Because what counts as the wrong thing to do differs from one class of problems to the next, there must be as many domain-specific subsystems as there are domains in which the definitions of successful behavioral outcomes are incommensurate.

(2) *Combinatorial explosion.* Combinatorial explosion paralyzes even moderately domain-general systems when encountering real-world complexity. As generality is increased by adding new dimensions to a problem space or new branch points to a decision tree, the computational load increases with catastrophic rapidity. A content-independent, specialization-free architecture contains no rules of relevance, procedural knowledge or privileged hypotheses, and so could not solve any biological problem of routine complexity in the amount of time an organism has to solve it (for discussion see, for example, Carey, 1985; Gallistel, Brown, Carey, Gelman, & Keil, 1991; Keil, 1989; Markman, 1989; Tooby & Cosmides, 1992). The question is not "How much specialization does a general purpose system require?" but rather "How many degrees of freedom can

a system *tolerate* – even a specialized, highly targeted one – and still compute decisions in useful, real-world time?" Combinatorics guarantee that real systems can only tolerate a small number. (Hence this problem cannot be solved by placing a few "constraints" on a general system.)

(3) *Clueless environments*. Content-free architectures are limited to knowing what can be validly derived by general processes from perceptual information. This sharply limits the range of problems they can solve: when the environment is clueless, the mechanism will be too. Domain-specific mechanisms are not limited in this way. They can be constructed to embody clues that fill in the blanks when perceptual evidence is lacking or difficult to obtain.

Consider the following adaptive problem. All plants foods contain an array of toxins. Ones that your liver metabolizes with ease sometimes harm a developing embryo. This subtle statistical relationship between the environment, eating behavior and fitness is ontogenetically "invisible": it cannot be observed or induced via general-purpose processes on the basis of perceptual evidence.[5] It can, however, be "observed" phylogenetically, by natural selection, because selection does not work by inference or simulation. Natural selection "counts up" the actual results of alternative designs operating in the real world, over millions of individuals, over thousands of generations, and weights these alternatives by the statistical distribution of their consequences: those design features that statistically lead to the best available outcome are retained. In this sense it is omniscient – it is not limited to what could be validly deduced by one individual, based on a short period of experience, it is not limited to what is locally perceivable, and it is not confused by spurious local correlations. As a result, it can build circuits – like those that regulate food choice during pregnancy – which embody privileged hypotheses that reflect and exploit these virtually unobservable relationships in the world. For example, the embryo/toxin problem is solved by a set of functionally specialized mechanisms that adjust the threshold on the mother's normal food aversion system (Profet, 1992). They lower it when the embryo is most at risk – thereby causing the food aversions, nausea and vomiting of early pregnancy – and raise it when caloric intake becomes a priority. As a result, the mother avoids ordinarily palatable foods when they would threaten the embryo: she responds adaptively to an ontogenetically invisible relationship. Functionally specialized designs allow organisms to solve a broad range of otherwise unsolvable adaptive problems. (For discussion of this design principle, see Cosmides & Tooby, 1987, 1994; Shepard, 1981, 1987; Tooby & Cosmides, 1990a.)

[5] Women ingest thousands of plant toxins every day; embryos self-abort for many reasons; early term abortions are often undetectable; the best trade-off between calories consumed and risk of teratogenesis is obscure.

In sum, architectures that do not come factory-equipped with sufficiently rich sets of content-specific machinery fail the solvability test. They could not have evolved, survived or propagated because they are incapable of solving even routine adaptive problems (Cosmides & Tooby, 1987, 1994; Tooby & Cosmides, 1992).

Natural selection, efficiency and functional specialization

Some researchers accept the conclusion that the human mind cannot consist solely of content-independent machinery, but nevertheless continue to believe that the mind needs very little content-specific organization to function. They believe that the preponderance of mental processes are content-independent and general purpose. Moreover, they believe that the correct null hypothesis – the parsimonious, prudent scientific stance – is to posit as few functionally specialized mechanisms as possible.

This stance ignores what is now known about the nature of the evolutionary process and the types of functional organization that it produces. Natural selection is a relentlessly hill-climbing process which tends to replace relatively less efficient designs with ones that perform better. Hence, in deciding which of two alternative designs is more likely to have evolved, their comparative performance on ancestral adaptive problems is the appropriate standard to use. Given this standard, positing a preponderance of general-purpose machinery is neither prudent nor parsimonious.[6] General-purpose mechanisms can't solve most adaptive problems at all, and in those few cases where one could, a specialized mechanism is likely to solve it more efficiently. The reason why is quite straightforward.

A general engineering principle is that the same machine is rarely capable of solving two different problems equally well. We have both cork-screws and cups because each solves a particular problem better than the other. It would be extremely difficult to open a bottle of wine with a cup or to drink from a cork-screw.

This same principle applies to the design of the human body. The heart is elegantly designed for pumping blood, but it is not good at detoxifying poisons; the liver is specialized for detoxifying poisons, but it cannot function as a pump. Pumping blood throughout the body and detoxifying poisons are two very different problems; consequently, the human body has a different machine for solving each of them. In biology, machines like these – ones that are specialized

[6] Parsimony applies to number of principles, not number of entities – physicists posit a small number of laws, not a small number of elements, molecules or stellar bodies. Epicycle upon epicycle would have to be added on to evolutionary theory to create a model in which less efficient designs frequently outcompeted more efficient ones.

and functionally distinct – are called *adaptive specializations* (Rozin, 1976). Specialization of design is natural selection's signature and its most common result (Williams, 1966).[7] In fact, the more important the adaptive problem, the more intensely natural selection tends to specialize and improve the performance of the mechanism for solving it.

There is no reason to believe that the human brain and mind are any exception. The cognitive programs that govern how you choose a mate should differ from those that govern how you choose your dinner. Different information-processing problems usually have different solutions. Implementing different solutions requires different, functionally distinct mechanisms (Sherry & Schacter, 1987). Speed, reliability and efficiency can be engineered into specialized mechanisms, because they do not need to engineer a compromise between mutually incompatible task demands: a jack of all trades – assuming one is possible at all – is necessarily a master of none. For this reason, one should expect the evolved architecture of the human mind to include many functionally distinct *cognitive* adaptive specializations.

And it does. For example, the learning mechanisms that govern language acquisition are different from those that govern the acquisition of food aversions, and both of these are different from the learning mechanisms that govern the acquisition of snake phobias (e.g., Cook, Hodes, & Lang, 1986; Cook & Mineka, 1989; Garcia, 1990; Mineka & Cook, 1988; Pinker, 1994; Ohman, Dimberg, & Ost, 1985; Ohman, Eriksson, & Olofsson, 1975). These adaptive specializations are *domain-specific*: the specialized design features that make them good at solving the problems that arise in one domain (avoiding venomous snakes) make them bad at solving the problems that arise in another (inducing a grammar). They are also *content-dependent*: they are activated by different kinds of content (speech versus screams), and their procedures are designed to accept different kinds of content as input (sentences versus snakes). A mind that applied relatively general-purpose reasoning circuits to all these problems, regardless of their content, would be a very clumsy problem-solver. But flexibility and efficiency of thought and action can be achieved by a mind that contains a battery of

[7]There are strict standards of evidence that must be met before a design feature can be considered an adaptation for performing function X. (1) The design feature must be species-typical; (2) function X must be an *adaptive* problem (i.e., a cross-generationally recurrent problem whose solution would have promoted the design feature's own reproduction); (3) the design feature must reliably develop (in the appropriate morphs) given the developmental circumstances that characterized its environment of evolutionary adaptedness; and, most importantly, (4) it must be shown that the design feature is particularly *well designed for performing function X*, and that it cannot be better explained as a by-product of some other adaptation or physical law. Contrary to popular belief, the following forms of "evidence" are *not* relevant: (1) showing that the design feature has a high heritability; (2) showing that variations in the environment do not affect its development; (3) showing that "learning" plays no role in its development. (Criteria for frequency-dependent adaptations differ. For refinements and complications, see Dawkins, 1982, 1986; Symons, 1992; Tooby & Cosmides, 1990b, 1992; and, especially, Williams, 1966, 1985).

special-purpose circuits. The mind is probably more like a Swiss army knife than an all-purpose blade: competent in so many situations because it has a large number of components – bottle opener, cork-screw, knife, toothpick, scissors – each of which is well designed for solving a different problem.

The functional architecture of the mind was designed by natural selection; natural selection is a hill-climbing process which produces mechanisms that solve adaptive problems well; a specialized design is usually able to solve a problem better than a more generalized one. It is unlikely that a process with these properties would design central processes that are general purpose and content-free. Consequently, one's default assumption should be that the architecture of the human mind is saturated with adaptive specializations.

How to find a needle in a haystack

The human brain is the most complex system scientists have ever tried to understand; identifying its components is enormously difficult. The more functionally integrated circuits it contains, the more difficult it will be to isolate and map any one of them. Looking for a functionally integrated mechanism within a multimodular mind is like looking for a needle in a haystack. The odds you'll find one are low unless you can radically narrow the search space. Marr's central insight was that you could do this by developing computational theories of the problems these mechanisms were designed to solve – for the human brain, the adaptive problems our hunter-gatherer ancestors faced.

The only behavioral scientists who still derive their hypotheses from intuition and folk psychology, rather than an evolutionarily based theory, are those who study humans.[8] The empirical advantages of using evolutionary biology to develop computational theories of adaptive problems have already been amply demonstrated in the study of non-human minds (e.g., Gallistel, 1990; Gould, 1982; Krebs & Davies, 1987; Real, 1991). We wanted to demonstrate its utility in studying the human mind. We thought an effective way of doing this would be to use an evolutionarily derived computational theory to discover cognitive mechanisms whose existence no one had previously suspected. Because most cognitive scientists still think of central processes as content-independent, we thought it would be particularly interesting to demonstrate the existence of central processes that are functionally specialized and content-dependent: domain-specific reasoning mechanisms.

Toward this end, we have conducted an experimental research program over the last 10 years, exploring the hypothesis that the human mind contains

[8]For a detailed analysis of the common arguments against the application of evolutionary biology to the study of the human mind, see Tooby and Cosmides (1992).

specialized circuits designed for reasoning about adaptive problems posed by the social world of our ancestors: social exchange, threat, coalitional action, mate choice, and so on. We initially focused on social exchange because (1) the evolutionary theory is clear and well developed, (2) the relevant selection pressures are strong, (3) paleoanthropological evidence suggests that hominids have been engaging in it for millions of years – more than enough time for selection to shape specialized mechanisms – and (4) humans in all cultures engage in social exchange. By starting with an adaptive problem hunter-gatherers are known to have faced, we could proceed to design experiments to test for associated cognitive specializations.

The evolutionary analysis of social exchange parallels the economist's concept of trade. Sometimes known as "reciprocal altruism", social exchange is an "I'll scratch your back if you scratch mine" principle (for evolutionary analyses see, for example, Axelrod, 1984; Axelrod & Hamilton, 1981; Boyd, 1988; Trivers, 1971; Williams, 1966.). Using evolvability constraints that biologists had already identified (some involving the Prisoners' Dilemma), we developed a computational theory of the information-processing problems that arise in this domain (Cosmides, 1985; Cosmides & Tooby, 1989). This gave us a principled basis for generating detailed hypotheses about the design of the circuits that generate social exchange in humans. Some of the design features we predicted are listed in Table 4.

For example, mathematical analyses had established cheater detection as a crucial adaptive problem. Circuits that generate social exchange will be selected out unless they allow individuals to detect those who fail to reciprocate favors – cheaters. This evolvability constraint led us directly to the hypothesis that humans might have evolved inference procedures that are specialized for detecting cheaters. We tested this hypothesis using the Wason selection task, which had originally been developed as a test of logical reasoning (Wason, 1966; Wason & Johnson-Laird, 1972).

A large literature already existed showing that people are not very good at detecting logical violations of "if–then" rules in Wason selection tasks, even when these rules deal with familiar content drawn from everyday life (e.g., Manktelow & Evans, 1979; Wason, 1983). For example, suppose you are skeptical when an astrologer tells you, "If a person is a Leo, then that person is brave," and you want to prove him wrong. In looking for exceptions to this rule, you will probably investigate people who you know are Leos, to see whether they are brave. Many people also have the impulse to investigate people who are brave, to see if they are Leos. Yet investigating brave people would be a waste of time; the astrologer said that all Leos are brave – not that all brave people are Leos – so finding a brave Virgo would prove nothing. And, if you are like most people, you probably won't realize that you need to investigate cowards. Yet a coward who turns out to be a Leo would represent a violation of the rule.

Table 4. *Reasoning about social exchange: evidence of special design*[a]

(a) The following design features were predicted and found	(b) The following by-product hypotheses were empirically eliminated
1. The algorithms governing reasoning about social contracts operate even in unfamiliar situations.	1. Familiarity cannot explain the social contract effect.
2. The definition of cheating that they embody depends on one's perspective.	2. It is not the case that social contract content merely facilitates the application of the rules of inference of the propositional calculus.
3. 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.	3. Social contract content does not merely "afford" clear thinking.
4. They embody implicational procedures specified by the computational theory.	4. Permission schema theory cannot explain the social contract effect; in other words, application of a generalized deontic logic cannot explain the results.
5. They include inference procedures specialized for cheater detection.	5. It is not the case that any problem involving payoffs will elicit the detection of violations.
6. Their cheater detection procedures cannot detect violations of social contracts that do not correspond to cheating.	
7. They do not include altruist detection procedures.	
8. They cannot operate so as to detect cheaters unless the rule has been assigned the cost–benefit representation of a social contract.	

[a]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.

If your mind had reasoning circuits specialized for detecting logical violations of rules, if would be immediately obvious to you that you should investigate Leos and cowards. But it is not intuitively obvious to most subjects. In general, fewer than 10% of subjects spontaneously realize this. Despite claims for the power of culture and "learning", even formal training in logical reasoning does little to boost performance (e.g., Cheng, Holyoak, Nisbett, & Oliver, 1986; Wason & Johnson-Laird, 1972).

However, we found that people who ordinarily cannot detect violations of "if–then" rules can do so easily and accurately when that violation represents cheating in a situation of social exchange. This is a situation in which one is entitled to a benefit only if one has fulfilled a requirement (e.g., "If you are to eat these cookies, then you must first fix your bed" or "If you are to eat cassava root, then you must have a tattoo on your face"). In these situations, the adaptively correct answer is immediately obvious to almost all subjects, who commonly experience a "pop out" effect. No formal training is needed. Whenever the content of a problem asks subjects to look for cheaters on a social exchange –even when the situation described is culturally unfamiliar and even bizarre –subjects experience the problem as simple to solve, and their performance jumps dramatically. Seventy to 90% of subjects get it right, the highest performance ever found for a task of this kind.

From a domain-general, formal view, investigating people eating cassava root and people without tattoos is logically equivalent to investigating Leos and cowards. But everywhere it has been tested, people do not treat social exchange problems as equivalent to other kinds of reasoning problems. Their minds distinguish social exchange contents, and apply domain-specific, content-dependent rules of inference that are adaptively appropriate only to that task. (For a review of the relevant experiments, see Cosmides & Tooby, 1992. For more detailed descriptions, see Cosmides, 1985, 1989; Cosmides & Tooby, 1989; Gigerenzer & Hug, 1992.)

We think that the goal of cognitive research should be to recover, out of carefully designed experimental studies, high-resolution "maps" of the intricate mechanisms that collectively constitute the cognitive architecture. Our evolutionarily derived computational theory of social exchange allowed us to construct experiments capable of detecting, isolating and mapping out previously unknown cognitive procedures. 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 (Cosmides & Tooby, 1989). Experimental tests have confirmed the presence of all the predicted design features that have been tested for so far. Those design features that have been tested and confirmed are listed in Table 4, 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 experiments like these have generated. The data seem best explained by the hypothesis that humans reliably develop circuits that are complexly specialized for reasoning about reciprocal social interactions.

Parallel lines of investigation have already identified two other domain-specialized reasoning mechanisms: one for reasoning about aggressive threats and one for reasoning about protection from hazards (e.g., Manktelow & Over, 1990; Tooby & Cosmides, 1989). We are now designing clinical tests to identify 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 that our various computational theories suggest.

Reasoning instincts

In our view, a large range of reasoning problems (like the astrological one) are difficult because (1) their content is not drawn from a domain for which humans evolved functionally specialized reasoning circuits, and (2) we lack the content-independent circuits necessary for performing certain logical operations ("logical reasoning"). In contrast, social exchange problems are easy because we do have evolved circuits specialized for reasoning about that important, evolutionarily long-enduring problem in social cognition. The inferences necessary for detecting cheaters are obvious to humans for the same reason that the inferences necessary for echolocation are obvious to a bat.

Instincts are often thought of as the polar opposite of reasoning. Non-human animals are widely believed to act through "instinct", while humans "gave up instincts" to become "the rational animal". But the reasoning circuits we have been investigating are complexly structured for solving a specific type of adaptive problem, they reliably develop in all normal human beings, they develop without any conscious effort and in the absence of any formal instruction, they are applied without any conscious awareness of their underlying logic, and they are distinct from more general abilities to process information or to behave intelligently. In other words, they have all the hallmarks of what one usually thinks of as an "instinct" (Pinker, 1994). Consequently, one can think of these specialized circuits as *reasoning instincts*. They make certain kinds of inferences just as easy, effortless and "natural" to us as humans, as spinning a web is to a spider or dead-reckoning is to a desert ant.

Three decades of research in cognitive psychology, evolutionary biology and neuroscience have shown that the central premise of the SSSM – that the mind is general purpose and content-free – is fundamentally misconceived. An alternative framework – sometimes called evolutionary psychology – is beginning to replace it (Tooby & Cosmides, 1992). According to this view, the evolved architecture of the human mind is full of specialized reasoning circuits and regulatory mechanisms that organize the way we interpret experience, construct knowledge and make decisions. These circuits inject certain recurrent concepts and motivations into our mental life, and they provide universal frames of meaning that allow us to understand the actions and intentions of others. Beneath the level of surface variability, all humans share certain views and assumptions about the nature of the world and human action by virtue of these universal reasoning circuits (Atran, 1990; Boyer, 1994; Brown, 1991; Carey & Gelman, 1991; Gelman & Hirschfeld, 1994; Keil, 1989; Leslie, 1987; Markman, 1990; Spelke, 1990; Sperber, 1985, 1990, 1994; Symons, 1979; Tooby & Cosmides, 1992).

III. Intuition is a misleading source of hypotheses because functionally specialized mechanisms create "instinct blindness"; computational theories are lenses that correct for instinct blindness

Intuitions about cognition: the limitations of an atheoretical approach

The adaptationist view of a multimodular mind was common at the turn of the century. Early experimental psychologists, such as William James and William McDougall, thought the mind is a collection of "faculties" or "instincts" that direct learning, reasoning and action (James, 1890; McDougall, 1908). These faculties were thought to embody sophisticated information-processing procedures that were domain-specific. In James's view, human behavior is so much more flexibly intelligent than that of other animals because we have *more* instincts than they do – not fewer (James, 1890).

The vocabulary may be archaic, but the model is modern. With every new discovery, it becomes more apparent that the evolved architecture of the human mind is densely multimodular – that it consists of an enormous collection of circuits, each specialized for performing a particular adaptive function. The study of perception and language has provided the most conspicuous examples, but evidence for the existence of learning instincts (Marler, 1991) and reasoning instincts is pouring in from all corners of the cognitive sciences (for examples, see Atran, 1990; Barkow, Cosmides, & Tooby, 1992; Baron-Cohen, Leslie, & Frith, 1985; A. Brown, 1990; D.E. Brown, 1991; Carey & Gelman, 1991; Cosmides & Tooby, in press; Daly & Wilson, 1988, 1994; Frith, 1989; Gelman & Hirschfeld, 1994; Gigerenzer, Hoffrage, & Kleinbolting, 1991; Leslie, 1988; Pinker, 1994; Rozin, 1976; Spelke, 1988; Sperber, 1994; Symons, 1979; Wilson & Daly, 1992; Wynn, 1992).

In spite of this consistent pattern, however, most cognitive scientists balk at the model of a brain crowded with specialized inference engines. Even Fodor, who has championed the case for modular processes, takes the traditional view that "central" processes are general purpose (Fodor, 1983). The notion that learning and reasoning are like perception and language – the complex product of a large collection of functionally specialized circuits – is deeply at war with our intuitions.

But so is the inherent indeterminacy in the position of electrons. It is uncomfortable but scientifically necessary to accept that common sense is the faculty that tells us the world is flat.[9] Our intuitions may feel authoritative and irresistibly compelling, and they may lead us to dismiss many ideas as ridiculous. But they are, nevertheless, an untrustworthy guide to the reality of subatomic particles or the evolved structure of the human mind.

In the case of central processes, we think human intuition is not merely untrustworthy: it is systematically misleading. Well-designed reasoning instincts should be invisible to our intuitions, even as they generate them – no more accessible to consciousness than retinal cells and line detectors, but just as important in creating our perception of the world.

Intuitively, we are all naive realists, experiencing the world as already parsed into objects, relationships, goals, foods, dangers, humans, words, sentences, social groups, motives, artifacts, animals, smiles, glares, relevances and saliences, the known and the obvious. This automatically manufactured universe, input as toy worlds into computers, seems like it could almost be tractable by that perennially elusive collection of general-purpose algorithms cognitive scientists keep expecting to find. But to produce this simplified world that we effortlessly experience, a vast sea of computational problems are being silently solved, out of awareness, by a host of functionally integrated circuits. These reasoning instincts are powerful inference engines, whose automatic, non-conscious operation creates our seamless experience of the world. The sense of clarity and self-evidence they generate is so potent it is difficult to see that the computational problems they solve even exist. As a result, we incorrectly locate the computationally manufactured simplicity that we experience as a natural property of the external world –as the pristine state of nature, not requiring any explanation or research.

Thus the "naturalness" of certain inferences acts to obstruct the discovery of the mechanisms that produced them. Cognitive instincts create problems for cognitive scientists. Precisely because they work so well – because they process information so effortlessly and automatically – we tend to be blind to their existence. Not suspecting they exist, we do not conduct research programs to find them.

To see that they exist, you need to envision an alternative conceptual universe. But these dedicated circuits structure our thought so powerfully that it can be difficult to imagine how things could be otherwise. As William James wrote:

> It takes . . . a mind debauched by learning to carry the process of making the natural seem strange, so far as to ask for the *why* of any instinctive human act. To the metaphysician alone can such questions occur as: why do we smile, when pleased, and not scowl? Why are we unable to talk to a

[9]This should not be surprising. Our intuitions were designed to generate adaptive behavior in Pleistocene hunter-gatherers, not useful theories for physicists and cognitive scientists.

crowd as we talk to a single friend? Why does a particular maiden turn our wits so upside-down? The common man can only say, *Of course* we smile, *of course* our heart palpitates at the sight of the crowd, *of course* we love the maiden, that beautiful soul clad in that perfect form, so palpably and flagrantly made for all eternity to be loved!

And so, probably, does each animal feel about the particular things it tends to do in the presence of particular objects. . . . To the lion it is the lioness which is made to be loved; to the bear, the she-bear. To the broody hen the notion would probably seems monstrous that there should be a creature in the world to whom a nestful of eggs was not the utterly fascinating and precious and never-to-be-too-much-sat-upon object which it is to her. (James, 1890)

For exactly this reason, intuition is an unreliable guide to points of interest in the human mind. Functionally specialized reasoning circuits will make certain inferences intuitive – so "natural" that there doesn't seem to be any phenomenon that is in need of explanation.

Consider, for example, sentences (1) and (2):

(1) If he's the victim of an unlucky tragedy, then we should pitch in to help him out.
(2) If he spends his time loafing and living off of others, then he doesn't deserve our help.

The inferences they express seem perfectly natural; there seems to be nothing to explain. They may not always be applicable, but they are perfectly intelligible.

But consider sentences (3) and (4):

*(3) If he's the victim of an unlucky tragedy, then he doesn't deserve our help.
*(4) If he spends his time loafing and living off of others, then we should pitch in to help him out.

Sentences (3) and (4) sound eccentric in a way that (1) and (2) do not. Yet they involve no *logical* contradictions. The inferences they embody seem to violate a grammar of social reasoning – in much the same way that "Alice might slowly" violates the grammar of English but "Alice might come" does not (Cosmides, 1985; Cosmides & Tooby, 1989, 1992). If so, then one needs to look for a reasoning device that can reliably generate (1) and (2) without also generating (3) and (4).

Realizing that *not* generating (3) and (4) is a design feature of the mechanism is tricky, however. Precisely because the device in question does not spontaneously generate inferences like (3) and (4), we rarely notice their absence or feel the need to explain it. And that is the root of the problem. There is a complex pattern to the inferences we generate, but seeing it requires a contrast between figure and ground; the geometry of a snowflake disappears against a white background. "Unnatural" inferences form the high contrast background necessary to see the complex geometry of the inferences that we do spontaneously generate. Yet these "unnatural" inferences are exactly the ones we don't produce. Without this background, the pattern can't be seen. As a result, we look neither for the pattern, nor for the mechanisms that generate it. And no one guesses that our central processes instantiate domain-specific grammars every bit as rich as that of a natural language (for more examples, see Table 5).

Hidden grammars

In the study of language, a grammar is defined as a finite set of rules that is capable of generating all the sentences of a language without generating any non-sentences; a sentence is defined as a string of words that members of a linguistic community would judge as well formed. In the study of reasoning, a grammar is a finite set of rules that can generate all appropriate inferences while not simultaneously generating inappropriate ones. If it is a grammar of social reasoning, then these inferences are about the domain of social motivation and

Table 5. *Inferences that violate a grammar of social reasoning*

(a)
I want to help him because he has helped me so often in the past.
I don't want to help him because whenever I'm in trouble he refuses to help me.

*I want to help him because whenever I'm in trouble he refuses to help me.
*I don't want to help him because he has helped me so often in the past.

(b)
I love my daughter. If you hurt her, I'll kill you.
*I love my daughter. If you hurt her, I'll kiss you.

(c)
If I help you now, then you must promise to help me.
*If I help you now, then you must promise to never help me.

(d)
He gave her something expecting nothing in return; she was touched.
*He gave her something expecting nothing in return; she was enraged.

(e)
She paid $5 for the book because the book was more valuable to her than $5.
*She paid $5 for the book because the book was less valuable to her than $5.

behavior; an "inappropriate" inference is defined as one that members of a social community would judge as incomprehensible or nonsensical.[10]

The cornerstone of any computational theory of the problem of language acquisition is the specification of a grammar. Discovering the grammar of a human language is so difficult, however, that there is an entire field – linguistics – devoted to the task. The task is difficult precisely because our linguistic inferences are generated by a "language instinct" (Pinker, 1994). One thing this set of specialized circuits can do is distinguish grammatical from ungrammatical sentences. But the rules that generate sentences – the grammar itself – operate effortlessly and automatically, hidden from our conscious awareness. Indeed, these complex rules are so opaque that just 40 years ago most linguists thought each human language – English, Chinese, Setswana – had a completely different grammar. Only recently have these grammars been recognized as minor variants on a Universal Grammar (UG): an invariant set of rules embodied in the brains of all human beings who are not neurologically impaired (Chomsky, 1980; Pinker, 1994).[11]

Universal grammars of social reasoning are invisible to cognitive scientists now for the same reason that UG was invisible to linguists for such a long time. The fact that the internal operations of the computational machinery in question are automatic and unconscious is a contributing factor; but the causes of invisibility go even deeper.

[10]The similarities between a grammar of language and a grammar of social reasoning run even deeper. Context can make a seemingly ungrammatical sentence grammatical. To pick a standard linguistic example, "The horse raced past the barn fell" seems ungrammatical when "raced" is categorized as the main verb of the sentence, but grammatical if the context indicates that there are two horses. "Fell" is then *recategorized* as the main verb, and "raced" as a passive verb within a prepositional phrase. Context can have the same effect on statements that seem socially ungrammatical. "I'll give you $1000 for your gum wrapper" seems eccentric – ungrammatical – because gum wrappers are considered worthless. It violates a grammatical constraint of social contract theory: that (benefit to offerer) > (cost to offerer) (Cosmides & Tooby, 1989). To become grammatical, the context must cause the violated constraint to be satisfied. For example, recategorizing the gum wrapper as something extremely valuable (potentially justifying the $1000 payment) would do this: the statement seems sensible if you are told that the speaker is a spy who knows the gum wrapper has a microdot with the key for breaking an enemy code.

[11]The term "innate" means different things to different scientific communities, but no person who uses the term means "immune to every environmental perturbation". UG is innate in the following sense: its intricate internal organization is the product of our species' genetic endowment in the same way that the internal organization of the eye is. Its neurological development is buffered against most naturally occurring variations in the physical and social environment. Certain environmental conditions are necessary to trigger the development of UG, but these conditions are not the source of its internal organization. As a result, all normal human beings raised in reasonably normal environments develop the same UG (e.g., Pinker, 1994). For an extensive discussion of how natural selection structures the relationships among genotype, phenotype and environment in development, see Tooby and Cosmides (1992).

Instinct blindness

UG is a small corner of hypothesis space; there are an indefinitely large number of grammars that are *not* variants of UG. To explain the fact that all natural languages fall within the bounds of UG, one must first realize that UG exists. To realize that it exists, one must realize that there are alternative grammars.

But this last step is where our imagination stumbles. The language instinct structures our thought so powerfully that alternative grammars are difficult to imagine. This is not an incidental feature of the language instinct; it is the language acquisition device's (LAD) principal adaptive function.[12] Any set of utterances a child hears is consistent with an infinite number of possible grammars, but only one of them is the grammar of its native language. A content-free learning mechanism would be forever lost in hypothesis space. The LAD is an adaptation to combinatorial explosion: by restricting the child's grammatical imagination to a very small subset of hypothesis space – hypotheses consistent with the principles of UG – it makes language acquisition possible. Its function is to generate grammatical inferences consistent with UG *without simultaneously* generating inconsistent ones. To do this, the LAD's structure must make alternative grammars literally unimaginable (at least by the language faculty).

This is good for the child learning language, but bad for the cognitive scientist, who needs to imagine these unimaginable grammars. Forming the plural through mirror reversal – so that the plural of "cat" is "tac" – is a rule in an alternative grammar. No child considers this possibility; the LAD cannot generate this rule. The cognitive scientist needs to know this, however, in order to characterize UG and produce a correct theory of the LAD's cognitive structure. UG is *what*, an algorithm is *how*. A proposed algorithm can be ruled out, for example, if formal analyses reveal that it produces both the mirror reverse rule and the "add 's' to a stem" rule.

Alternative grammars – and hence Universal Grammar – were difficult to discover because circuits designed to generate only a small subset of all grammatical inferences in the child also do so in the linguist. This property of the language instinct is crucial to its adaptive function. But it caused a form of theoretical blindness in linguists, which obstructed the discovery of UG and of the language instinct itself. One can think of this phenomenon as *instinct blindness*.

Discovering a grammar of social reasoning is likely to prove just as difficult as discovering the grammar of a language, and for exactly the same reasons. Yet

[12] As a side-effect, it can also solve problems that played no causal role in its selective history. For example, the LAD was not designed to support writing, but its properties made the design and spread of this cultural invention possible.

there is no field, parallel to linguistics, that is devoted to this task; indeed, very few individuals even recognize the need for such a grammar, let alone such a field (for exceptions, see Cosmides, 1985; Cosmides & Tooby, 1989, 1992; Fiske, 1991; Jackendoff, 1992).

Our intuitions blind us not only to the existence of instincts, but to their complexity. The phenomenal experience of an activity as "easy" or "natural" often leads scientists to assume that the processes that give rise to it are simple. Legend has it that in the early days of artificial intelligence, Marvin Minsky assigned the development of machine vision to a graduate student as a summer project. This illusion of simplicity hampered vision research for years:

> ... in the 1960s almost no one realized that machine vision was difficult. The field had to go through [a series of fiascoes] before it was at last realized that here were some problems that had to be taken seriously. The reason for this misperception is that we humans are ourselves so good at vision. (Marr, 1982, p. 16)

Phenomenally, seeing seems simple. It is effortless, automatic, reliable, fast, unconscious and requires no explicit instruction. But seeing is effortless, automatic, reliable, fast, and unconscious precisely because there is a vast array of complex, dedicated computational machinery that makes this possible.

Most cognitive scientists don't realize it, but they are grossly underestimating the complexity of our central processes. To find someone beautiful, to fall in love, to feel jealous, to experience moral outrage, to fear disease, to reciprocate a favor, to initiate an attack, to deduce a tool's function from its shape – and a myriad other cognitive accomplishments – can seem as simple and automatic and effortless as opening your eyes and seeing. But this apparent simplicity is possible only because there is a vast array of complex computational machinery supporting and regulating these activities. The human cognitive architecture probably embodies a large number of domain-specific "grammars", targeting not just the domain of social life, but also disease, botany, tool-making, animal behavior, foraging and many other situations that our hunter-gatherer ancestors had to cope with on a regular basis.

Research on the computational machinery responsible for these kinds of inferences, choices and preferences – especially the social ones – is almost totally absent in the cognitive sciences. This is a remarkable omission, from an evolutionary point of view. Instinct blindness is one culprit; extreme and unfounded claims about cultural relativity is another (e.g., Brown, 1991; Sperber, 1982; Tooby & Cosmides, 1992).

Anthropological malpractice

As a result of the rhetoric of anthropologists, most cognitive researchers have, as part of their standard intellectual furniture, a confidence that cultural relativity is an empirically established finding of wide applicability (see discussion of the Standard Social Science Model in Tooby & Cosmides, 1992). Consequently, most scientists harbor the incorrect impression that there is no "Universal Grammar" of social reasoning to be discovered. According to this view, a grammar of social reasoning might exist in each culture, but these grammars will differ dramatically and capriciously from one culture to the next. In its most extreme form, the relativist position holds that the grammars of different cultures are utterly incommensurate – that there is no transformation that can map the rules of one onto the rules of another. If so, then these rules cannot be expressions of an underlying UG of social reasoning.

Among anthropologists, however, cultural relativism is an interpretation imposed as an article of faith – not a conclusion based on scientific data (Brown, 1991; Sperber, 1982; Tooby & Cosmides, 1992).[13] Indeed, Maurice Bloch, a prominent member of the field, has complained that it is the "professional malpractice of anthropologists to exaggerate the exotic character of other cultures" (Bloch, 1977). To some degree, this is a self-legitimizing institutional pressure: why go long distances to study things that could be studied at home (Brown, 1991)? More importantly, however, anthropologists are just as oblivious to what is universally natural for the human mind as the rest of us. Their attention is drawn to what differs from culture to culture, not what is absent from all cultures or what differs from species to species. Drawing on their cognitive instincts, they understand, automatically and without reflection, much of what happens in other cultures. They know they can work out exchanges without language, or see a smile, a shared look, or an aggressive gesture and infer its meaning and its referent. Indeed, they operate within a huge set of implicit panhuman assumptions that allow them to decode the residue of human life that does differ from place to place (Sperber, 1982; Tooby & Cosmides, 1992).

The notion of universal human reasoning instincts – including social reasoning instincts – is completely compatible with the ethnographic record. It is more than empirically reasonable; it is a logical necessity, for the reasons discussed above. Indeed, without universal reasoning instincts, the acquisition of one's "culture" would be literally impossible, because one wouldn't be able to infer which representations, out of the infinite universe of possibilities, existed in the minds of other members of the culture (Boyer, 1994; Chomsky, 1980; Sperber, 1985, 1990; Tooby & Cosmides, 1992).

Instinct blindness is a side-effect of any instinct whose function is to generate some inferences or behaviors without simultaneously generating others. This is a

[13]For a history and discussion of how unsupported relativist claims gained widespread acceptance in the social sciences, see Brown (1991) and Tooby and Cosmides (1992).

very general property of instincts, because combinatorial explosion is a very general selection pressure (for discussion, see Tooby & Cosmides, 1992). The fact that human instincts are difficult for human minds to discover is a side-effect of their adaptive function.

Many aspects of the human mind can't be seen by the naked "I" – by intuition unaided by theory. A good theory rips away the veil of naturalness and familiarity that our own minds create, exposing computational problems whose existence we never even imagined. The cognitive sciences need theoretical guidance that is grounded in something beyond intuition. Otherwise, we're flying blind.

Corrective lenses

There are various ways of overcoming instinct blindness. One of the most common is the study of non-human minds that differ profoundly from our own – animal minds and electronic minds, broody hens and AI programs. Linguists were awakened to the existence of alternative grammars by the creation of computer "languages", which are not variants of UG. These languages "made the natural seem strange", inspiring linguists to generate even stranger grammars. To do this, they had to escape the confines of their intuitions, which they did through the use of mathematical logic and the theory of computation. In William James's terms, they debauched their minds with learning.

The study of animal behavior is another time-honored method for debauching the mind – the one used by William James himself. Hermaphroditic worms, colonies of ant sisters who come in three "genders" (sterile workers, soldiers, queens), male langur monkeys who commit systematic infanticide when they join a troop, flies who are attracted to the smell of dung, polyandrous jacanas who mate with a male after breaking the eggs he was incubating for a rival female, fish who change sex when the composition of their social group changes, female praying mantises who eat their mate's head while copulating with him –other animals engage in behaviors that truly are exotic by human standards. Human cultural variation is trivial in comparison. Observing behaviors caused by alternative instincts jars us into recognizing the specificity and multiplicity of our own instincts.

Observations like these tell us what we are not, but not what we are. That's why theoretical biology is so important. It provides positive theories of what kinds of cognitive programs we should expect to find in species that evolved under various ecological conditions: theories of what and why. Evolutionary biology's formal theories are powerful lenses that correct for instinct blindness. In their focus, the intricate outlines of the mind's design stand out in sharp relief.

References

Atran, S. (1990). *The cognitive foundations of natural history*. New York: Cambridge University Press.
Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.
Axelrod, R., & Hamilton, W.D. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.
Barkow, J., Cosmides, L., & Tooby, J. (Eds.) (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.
Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*, 37–46.
Bloch, M. (1977). The past and the present in the present. *Man*, *12*, 278–292.
Boyd, R. (1988). Is the repeated prisoner's dilemma a good model of reciprocal altruism? *Ethology and Sociobiology*, *9*, 211–222.
Boyer, P. (1994). *The naturalness of religious ideas*. Berkeley: University of California Press.
Brothers, L. (1990). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience*, *1*, 27–51.
Brown, A. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science*, *14*, 107–133.
Brown, D.E. (1991). *Human universals*. New York: McGraw-Hill.
Carey, S. (1985). Constraints on semantic development. In J. Mehler & R. Fox (Eds.), *Neonate cognition* (pp. 381–398). Hillsdale, NJ: Erlbaum.
Carey, S., & Gelman, R. (Eds.) (1991). *The epigenesis of mind*. Hillsdale, NJ: Erlbaum.
Cheney, D.L., & Seyfarth, R. (1990). *How monkeys see the world*. Chicago: University of Chicago Press.
Cheng, P., Holyoak, K., Nisbett, R., & Oliver, L. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, *18*, 293–328.
Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.
Clutton-Brock, T.H., & Harvey, P. (1979). Comparison and adaptation. *Proceedings of the Royal Society, London B*, *205*, 547–565.
Cook, E.W., III, Hodes, R.L., & Lang, P.J. (1986). Preparedness and phobia: Effects of stimulus content on human visceral conditioning. *Journal of Abnormal Psychology*, *95*, 195–207.
Cook, M., & Mineka, S. (1989). Observational conditioning of fear to fear-relevant versus fear-irrelevant stimuli in rhesus monkeys. *Journal of Abnormal Psychology*, *98*, 448–459.
Cosmides, L. (1985). Deduction or Darwinian algorithms? An explanation of the "elusive" content effect on the Wason selection task. Doctoral dissertation, Department of Psychology, 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., & Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupre (Ed.), *The latest on the best: Essays on evolution and optimality*. Cambridge, MA: MIT Press.
Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture, Part II. Case study: A computational theory of social exchange. *Ethology and Sociobiology*, *10*, 51–97.
Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.
Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In S. Gelman & L. Hirschfeld (Eds.), *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.
Cosmides, L., & Tooby, J. (in press). Are humans good intuitive statisticians after all? Rethinking some conclusion of the literature on judgment under uncertainty. *Cognition*.
Daly, M., & Wilson, M. (1983). *Sex, evolution and behavior*. Boston: Wadsworth.
Daly, M., & Wilson, M. (1988). *Homicide*. Hawthorne, NY: Aldine de Gruyter.
Daly, M., & Wilson, M. (1994). Discriminative parental solicitude and the relevance of evolutionary

models to the analysis of motivational systems. In M. Gazzaniga (Ed.), *The cognitive neurosciences*. Cambridge, MA: MIT Press.
Dawkins, R. (1982). *The extended phenotype*. Oxford: W.H. Freeman.
Dawkins, R. (1986). *The blind watchmaker*. New York: Norton.
de Waal, F.B.M., & Luttrell, L.M. (1988). Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethology and Sociobiology*, *9*, 101–118.
Fischer, E.A. (1988). Simultaneous hermaphroditism, tit-for-tat, and the evolutionary stability of social systems. *Ethology and Sociobiology*, *9*, 119–136.
Fiske, A.P. (1991). *Structures of social life: The four elementary forms of human relations*. New York: Free Press.
Fodor, J.A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
Frith, U. (1989). *Autism: Explaining the enigma*. Oxford: Blackwell.
Gallistel, C.R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
Gallistel, C.R., Brown, A.L., Carey, S., Gelman, R., & Keil, F.C. (1991). Lessons from animal learning for the study of cognitive development. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind*. Hillsdale, NJ: Erlbaum.
Garcia, J. (1990). Learning without memory. *Journal of Cognitive Neuroscience*, *2*, 287–305.
Gelman, S., & Hirschfeld, L. (Eds.) (1994). *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.
Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikean theory of confidence. *Psychological Review*, *98*, 506–528.
Gigerenzer, G., & Hug, K. (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.
Gould, S.J., & Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society, London B*, *205*, 581–598.
Jackendoff, R. (1992). *Languages of the mind*. Cambridge, MA: MIT Press.
James, W. (1890). *Principles of psychology*. New York: Henry Holt.
Keil, F.C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
Krebs, J.R., & Davies, N.B. (1987). *An introduction to behavioural ecology*. Oxford: Blackwell Scientific Publications.
Leslie, A.M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review*, *94*, 412–426.
Leslie, A.M. (1988). The necessity of illusion: Perception and thought in infancy. In L. Weiskrantz (Ed.), *Thought without language* (pp. 185–210). Oxford: Clarendon Press.
Manktelow, K.I., & Evans, J. St.B.T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology*, *70*, 477–488.
Manktelow, K.I., & Over, D.E. (1990). Deontic thought and the selection task. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie, & G. Erdos (Eds.), *Lines of thinking* (Vol. 1). Chichester: Wiley.
Markman, E.M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.
Markman, E. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*, 57–77.
Marler, P. (1991). The instinct to learn. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind*. Hillsdale, NJ: Erlbaum.
Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
Mayr, E. (1983). How to carry out the adaptationist program? *The American Naturalist*, *121*, 324–334.
McDougall, W. (1908/1916). *Introduction to social psychology*. Boston: John W. Luce.
Mineka, S., & Cook, M. (1988). Social learning and the acquisition of snake fear in monkeys. In T.R. Zentall & B.G. Galef (Eds.), *Social learning: Psychological and biological perspectives*. (pp. 51–73). Hillsdale, NJ: Erlbaum.
Ohman, A., Dimberg, U., & Ost, L.G. (1985). Biological constraints on the fear response. In S. Reiss & R. Bootsin (Eds.), *Theoretical issues in behavior therapy* (pp. 123–175). New York: Academic Press.

Ohman, A., Eriksson, A., & Olofsson, A. (1975). One-trial learning and superior resistance to extinction of autonomic responses conditioned to potentially phobic stimuli. *Journal of Comparative and Physiological Psychology*, *88*, 619–627.

Pinker, S. (1994). *The language instinct*. New York: Morrow.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*, 707–727.

Profet, M. (1992). Pregnancy sickness as adaptation: A deterrent to maternal ingestion of teratogens. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Pylyshyn, Z.W. (Ed.) (1987). *The robot's dilemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex.

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 J.M. Sprague & A.N. Epstein (Eds.), *Progress in psychobiology and physiological psychology*. New York: Academic Press.

Shepard, R.N. (1981). Psychophysical complementarity. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum.

Shepard, R.N. (1987). Evolution of a mesh between principles of the mind and regularities of the world. In J. Dupre (Ed.), *The latest on the best: Essays on evolution and optimality*. Cambridge, MA: MIT Press.

Sherry, D.F., & Schacter, D.L. (1987). The evolution of multiple memory systems. *Psychological Review*, *94*, 439–454.

Smuts, B. (1986). *Sex and friendship in baboons*. Hawthorne: Aldine.

Spelke, E.S. (1988). The origins of physical knowledge. In L. Weiskrantz (Ed.), *Thought without language* (pp. 168–184). Oxford: Clarendon Press.

Spelke, E. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.

Sperber, D. (1975) *Rethinking symbolism*, transl. Alice Morton. Cambridge, UK: Cambridge University Press.

Sperber, D. (1982). *On anthropological knowledge*. Cambridge, UK: Cambridge University Press.

Sperber, D. (1985). Anthropology and psychology: Towards an epidemiology of representations. *Man (N.S.)*, *20*, 73–89.

Sperber, D. (1990). The epidemiology of beliefs. In C. Fraser & G. Geskell (Eds.), *Psychological studies of widespread beliefs*. Oxford: Clarendon Press.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In S. Gelman & L. Hirschfeld, (Eds.), *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.

Symons, D. (1979). *The evolution of human sexuality*. New York: Oxford University Press.

Symons, D. (1992). On the use and misuse of Darwinism in the study of human behavior. In J. Barkow, L. Cosmides, & J. Tooby, (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Tooby, J., & Cosmides, L. (1989). The logic of threat: Evidence for another cognitive adaptation? Paper presented at the Human Behavior and Evolution Society, Evanston, IL.

Tooby, J., & Cosmides, L. (1990a). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, *11*, 375–424.

Tooby, J., & Cosmides, L. (1990b). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality*, *58*, 17–67.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Tooby J., & DeVore, I. (1987). The reconstruction of hominid behavioral evolution through strategic modeling. In W. Kinzey (Ed.), *Primate models of hominid behavior*. New York: SUNY Press.

Trivers, R.L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35–57.

Wason, P. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology*. Harmondsworth: Penguin.

Wason, P. (1983). Realism and rationality in the selection task. In J. St.B.T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.

Wason, P., & Johnson-Laird, P.N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.

Wilkinson, G.S. (1988). Reciprocal altruism in bats and other mammals. *Ethology and Sociobiology*, *9*, 85–100.

Wilkinson, G.S. (1990). Food sharing in vampire bats. *Scientific American*, 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. (1985). A defense of reductionism in evolutionary biology. *Oxford surveys in evolutionary biology*, *2*, 1–27.

Williams, G.C., & Nesse, R.M. (1991). The dawn of Darwinian medicine. *Quarterly Review of Biology*, *66*, 1–22.

Wilson, M., & Daly, M. (1992). The man who mistook his wife for a chattel. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*, 749–750.