From:

The Adapted Mind:
Evolutionary psychology and the
generation of culture

J. Barkow, L. Cosmides & J. Tooby, editors

1992. New York: Oxford University Press

3

Cognitive Adaptations for Social Exchange

LEDA COSMIDES AND JOHN TOOBY

INTRODUCTION

> 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

Research Background

The human mind is the most complex natural phenomenon humans have yet encountered, and Darwin's gift to those who wish to understand it is a knowledge of the process that created it and gave it its distinctive organization: evolution. Because we know that the human mind is the product of the evolutionary process, we know something vitally illuminating: that, aside from those properties acquired by chance, the mind consists of a set of adaptations, designed to solve the long-standing adaptive problems humans encountered as hunter-gatherers. Such a view is uncontroversial to most behavioral scientists when applied to topics such as vision or balance. Yet adaptationist approaches to human psychology are considered radical—or even transparently false—when applied to most other areas of human thought and action, especially social behavior. Nevertheless, the logic of the adaptationist postion is completely general, and a dispassionate evaluation of its implications leads to the expectation that humans should have evolved a constellation of cognitive adaptations to social life. Our ancestors have been members of social groups and engaging in social interactions for millions and probably tens of millions of years. To behave adaptively, they not only needed to construct a spatial map of the objects disclosed to them by their retinas, but a social map of the persons, relationships, motives, interactions, emotions, and intentions that made up their social world.

Our view, then, is that humans have a faculty of social cognition, consisting of a rich collection of dedicated, functionally specialized, interrelated modules (i.e., functionally isolable subunits, mechanisms, mental organs, etc.), organized to collectively guide thought and behavior with respect to the evolutionarily recurrent adaptive problems posed by the social world. Nonetheless, if such a view has merit, it not only must be argued for on theoretical grounds—however compelling—but also must be substantiated by experimental evidence, as well as by converging lines of empirical support drawn from related fields such as neuroscience, linguistics, and anthropology. The eventual goal is to recover out of carefully designed experimental studies "high-resolution" maps of the intricate mechanisms involved. Such an approach is intended to exploit the signal virtue of cognitive psychology: With its emphasis on mechanisms, cognitive approaches allow causal pathways to be precisely specified through reference to explicitly described algorithms and representations.

Toward this end, we have conducted an experimental research program over the last eight years, exploring the hypothesis that the human mind contains algorithms (specialized mechanisms) designed for reasoning about social exchange. The topic of reasoning about social exchange was selected for several reasons. In the first place, as we will discuss, many aspects of the evolutionary theory of social exchange (also sometimes called cooperation, reciprocal altruism, or reciprocation) are relatively well developed and unambiguous. Consequently, certain features of the functional logic of social exchange can be confidently relied on in constructing hypotheses about the structure of the information-processing procedures that this activity requires.

In the second place, complex adaptations are constructed in response to evolutionarily long-enduring problems, and it is likely that our ancestors have engaged in social exhange for at least several million years. Several converging lines of evidence support this view. Social exchange behavior is both universal and highly elaborated across all human cultures—including hunter-gatherer cultures (e.g., Cashdan, 1989; Lee & DeVore, 1968; Sharp, 1952; Wiessner, 1982)—as would be expected if it were an ancient and central part of human social life. If social exchange were merely a recent invention, like writing or rice cultivation, 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. Moreover, the nearest relatives to the hominid line, the chimpanzees, also engage in certain types of sophisticated reciprocation (de Waal, 1982; de Waal & Luttrell, 1988), which implies that some cognitive adaptations to social exchange were present in the hominid lineage at least as far back as the common ancestors that we share with the chimpanzees, five to ten million years ago. Finally, paleoanthropological evidence also supports the view that exchange behavior is extremely ancient (e.g., Isaac, 1978; McGrew & Feistner, this volume; Tooby & DeVore, 1987). These facts, plus the existence of reciprocation among members of primate species that are even more distantly related to us than chimpanzees—such as macaques and baboons (Packer, 1977; de Waal & Luttrell, 1988)—strongly support the view that situations involving social exchange have constituted a long-enduring selection pressure on hominids.

The third reason we selected reasoning about social exchange as the focus of this experimental series was that theories about reasoning and rationality have played a central role in both cognitive science and the social sciences. Research in this area can, as a result, function as a powerful test of certain traditional social science postulates. An adaptationist approach to human psychology is often viewed as radical or false not because of gaps in its logic or any comparative lack of evidence for its hypotheses, but because it violates certain privileged tenets of this century's dominant behavioral and social science paradigm—what we have called elsewhere the Standard Social Science Model (see Tooby & Cosmides, this volume). According to this view, all of the specific content of the human mind originally derives from the "outside"—from the environment and the social world—and the evolved architecture of the mind consists solely or predominantly of a small number of general-purpose mechanisms that are *content-independent*, and which sail under names such as "learning," "induction," "intelligence," "imitation," "rationality," "the capacity for culture," or, simply, "culture." On this view, the same mechanisms are thought to govern how one acquires a language and how one acquires a gender identity. This is because the mechanisms that govern reasoning, learning, and memory are assumed to operate uniformly across all domains: They do not impart content, they are not imbued with content, and they have no features specialized for processing particular kinds of content. Hypotheses that are inconsistent with this content-free view of the mind are, a priori, not considered credible, and the data that support them are usually explained away by invoking as alternatives the operation of general-purpose processes of an unspecified nature. Strong results indicating the involvement of domain-specific adaptations in areas such as perception, language, and emotion have sometimes—though grudgingly—been accepted as genuine, but have been ghettoized as exceptional cases, not characteristic of the great majority of mental processes.

In this dialogue, reasoning has served as the paradigm case of the "general-purpose" psychological process: It has been viewed as preeminently characteristic of those processes that are purportedly the central engine of the human mind. Even vigorous advocates of modularity have held so-called central processes, such as reasoning, to be general-purpose and content-independent (e.g., Fodor, 1983). Consequently, we felt that reasoning about social exchange offered an excellent opportunity to cut to the quick of the controversy. If even human reasoning, the doctrinal "citadel" of the advocates of content-free, general-purpose processes, turns out to include a large number of content-dependent cognitive adaptations, then the presumption that psychological mechanisms are characteristically domain-general and originally content-free can no longer be accorded privileged status. Such results would jeopardize the assumption that whenever content-dependent psychological phenomena are found, they necessarily imply the prior action of cultural or environmental shaping. Instead, such results would add credibility to the contrary view that the mind is richly textured with content-specialized psychological adaptations.

Evolutionary biologists have developed useful criteria for establishing the existence of adaptations (e.g., Dawkins 1982, 1986; Symons, this volume; Thornhill, 1991; Tooby & Cosmides, 1990b; Williams, 1966, 1985), and these crtieria are helpful in evaluating experimental evidence that bears on these two positions. Adaptations can be recognized by "evidence of special design" (Williams, 1966)—that is, by recognizing that features of the evolved species-typical design of an organism are "components of some special problem-solving machinery" that solves an evolutionarily long-standing problem (Williams, 1985, p. 1). Standards for recognizing special design include factors such as economy, efficiency, complexity, precision, specialization, and reliability, which—like a key fitting a lock—render the design too good a solution to an adaptive problem to have arisen by chance (Williams, 1966). For example, the eye is extremely well suited for the detection and extraction of information presented by ambient light, and poorly designed as an orifice for ingesting food or as armor to protect the vulnerable brain from sharp objects. It displays many properties that are only plausibly interpreted as design features for solving the problem of vision. Moreover, the properties of an adaptation can be used to identify the class of problems, at the correct level of specificity or generality, that the adaptation was designed to solve. The eye allows humans to see hyenas, but that does not mean it is an adaptation that evolved particularly for hyena detection: There are no features that render it better designed for seeing hyenas than for seeing any of a far larger class of comparable objects. These principles governing adpatations can be developed into a series of methods for empirically arbitrating the dispute between traditional and domain-specific views of the mind. The Standard Social Science Model and evolutionary psychological approaches differ most strongly on the grounds of functional specialization, of content-specificity, and of evolutionary appropriateness (Tooby & Cosmides, this volume).

According to the evolutionary psychological approach to social cognition outlined here and elsewhere (Cosmides, 1985, 1989; Cosmides & Tooby, 1987, 1989; Tooby, 1985; Tooby & Cosmides, 1989, 1990b), the mind should contain organized systems of inference that are specialized for solving various families of problem, such as social exchange, threat, coalitional relations, and mate choice. Advocates of evolutionary views do not deny that humans learn, reason, develop, or acquire a culture; however, they do argue that these functions are accomplished at least in part through the operation of cognitive mechanisms that are content-specialized—mechanisms that are activated by particular content domains and that are designed to process information from those domains. Each cognitive specialization is expected to contain design features targeted to mesh with the recurrent structure of its characteristic problem type, as encountered under Pleistocene conditions. Consequently, one expects cognitive adaptations specialized for reasoning about social exchange to have some design features that are particular and appropriate to social exchange, but that are not activated by or applied to other content domains.

In contrast, the Standard Social Science Model predicts that the reasoning procedures applied to situations of social exchange should be the same reasoning procedures that are applied to other kinds of content. On this view, reasoning is viewed as the operation of content-independent procedures, such as formal logic, applied impartially and uniformly to every problem, regardless of the nature of the content involved. There should be nothing in the evolved structure of the mind—no content-sensitive procedures, no special representational format—that is more appropriate for reasoning about social exchange than about hat racks, rutabagas, warfare, Hinayana scripture, turbulence, or textuality. In other words, the standard view is that the faculty of reasoning consists of a small number of processes that are designed to solve the most inclusive and general class of reasoning problems possible—a class not defined in terms of its content, as the class includes all potential contents equally. On this view, any variability in reasoning due to content must be the product of experiential variables such as familiarity or explicit instruction.

For these reasons, the questions of interest for this experimental program include the following: Do patterns of performance on problems that require reasoning about social exchange reflect content-general rules of logic? Do patterns of performance on social exchange content, as compared with other contents, show systematic differences? If so, can these differences be explained through invoking general-purpose variables such as familiarity? Does the complexly articulated performance of subjects on social exchange problems have the detailed properties predicted in advance by an evolutionary analysis of the design features required for a cognitive adaptation to social exchange? By answering these and related questions, building from one experimental result to the next, the functional structure of human cognitive adaptations for reasoning about social exchange can begin to be delineated, and the adequacy of the Standard Social Science Model can be assessed.

Standard Analyses of the Evolution of Altruism

Natural selection is a feedback process that is driven by the differential reproduction of alternative designs. If a change in an organism's design allows it to outreproduce the alternative designs in the population, then that design change will become more common—it will be *selected for.* If this reproductive advantage continues, then over many generations that design change will spread through the population until all members of the species have it. Design changes that enhance reproduction are selected for; those that hinder reproduction relative to others are selected against and, therefore, tend to disappear. This ongoing process leads over time to the accumulation of designs organized for reproduction.

Consider, then, a design change that appears to *decrease* the reproduction of an individual who has it while simultaneously increasing the reproduction of other individuals. How could such a design change possibly spread through the population? At first glance, it would seem that a design feature that had this property would be selected against.

Yet many organisms do engage in behaviors that decrease their own reproduction while enhancing that of others. One chimpanzee will endanger itself to help another in a fight (de Waal, 1982). A vampire bat will feed blood that it has collected from its prey to a hungry conspecific (Wilkinson, 1988, 1990). A ground squirrel will warn others of the presence of a predator by emitting an alarm call that can draw the predator's attention to itself (Sherman, 1977). Among many species of social insects, workers forgo reproduction entirely in order to help raise their sisters (Wilson, 1971). People sometimes put themselves at great peril to help their fellow human beings, and carry out innumerable acts on a daily basis whose purpose is to help others. If a psychological mechanism generates such behavior on a regular basis, how could it possibly have been selected for?

Evolutionary biologists call this the "problem of altruism." An "altruistic" design feature is an aspect of the phenotype that is designed to produce some effect that enhances the reproduction of other individuals even though it may cause the individual who has it to reproduce less. The question is, how can designs that generate such behavior spread through a population until they become universal and species-typical?

So far, evolutionary biologists have provided two answers to the problem of altruism. The first. kin selection theory (or inclusive fitness theory), was proposed by W. D. Hamilton in 1964 (see also Maynard Smith, 1964; Williams & Williams, 1957). Imagine a design change that causes an individual to increase the reproduction of that individual's relatives, but that decreases the individual's own reproduction. There is some probability, *r*, that the kin member who receives the help has inherited that very same design change from a common ancestor. Therefore, the design change—through helping the relative to reproduce—may be spreading new copies of itself in the population, even though it is simultaneously decreasing the rate at which it creates new copies of itself through the individual it is in, by slowing the reproduction of that particular individual. Whenever a design change affects both direct reproduction and kin reproduction, there is a trade-off between these two different avenues by which a design change can be reproduced. The fate of the design change will be determined by how much it helps (or harms) the relative, how much it harms (or helps) the helper, and the probability the relative shares the design change by virtue of their sharing common ancestors. By using what was, in effect, mathematical game theory, Hamilton showed that a "helping design" can spread through the population *if* it causes an organism to help a kin member whenever the cost to the organism's own reproduction is offset by the benefit to the reproduction of its kin member, discounted by the probability, *r*, that the kin member has inherited the same helping design. Although helping under these circumstances decreases the helper's *personal* reproduction, through its effect on other individuals it causes a net increase in the reproduction of the helping design itself in the population.

Consequently, if C_i and B_i refer to costs and benefits to an individual *i*'s own reproduction, then an altruistic design change can be selected for if it causes *i* to help *j* whenever $C_i < r_{ij}B_j$. Any design change that causes an individual to help more than this—or less than this—would be selected against. This constraint is completely general and falls out of the logic of natural selection theory: It should be true of any species on any planet at any time. A species may be solitary, and individuals may have no social interactions with their relatives; but if members of a species consistently interact socially with their relatives in ways that affect their reproduction, then they will be selected to evolve information-processing mechanisms that produce behavior that respects this constraint.

Because it suggested a rich set of hypotheses about phenotypic design, kin selection theory allowed animal behavior researchers to discover a flood of new phenomena. They began to find that the altruistic behavior of many species shows the design features that one would expect if their information-processing mechanisms had been shaped by kin selection. For example, ground squirrels are far more likely to give an alarm call *if* a close relative lives nearby (Sherman, 1977), and they have psychological mechanisms that allow them to discriminate full siblings from half siblings from unrelated individuals (Hanken & Sherman, 1981; Holmes & Sherman, 1982). Similarly, kinship is a major predictor of whether a vampire bat will share its food with a particular individual (Wilkinson, 1988, 1990). Most strikingly, kin selection theory (e.g., Hamilton, 1964; Williams & Williams, 1957) finally explained the existence of the sterile worker castes in the eusocial insects that had so troubled Darwin, providing an elegant set of hypotheses concerning how eusocial insects should allocate their reproductive effort among sisters, half-sisters, brothers and offspring, which have since been tested and confirmed (e.g., Frumhoff & Baker, 1988; Frumhoff & Schneider, 1987; Trivers & Hare, 1976).

The realization that a design feature can make copies of itself not only by affecting the reproductive success of its bearer, but also by affecting the reproductive success of its bearers' kin, led to a new definition of the concept of fitness. Previously, evolutionary biologists spoke of a design's "Darwinian fitness": its effect on the number of offspring produced by an individual who has the design. But since Hamilton, one speaks of a design's "inclusive fitness": its effect on the number of offspring produced by an individual who has the design *plus* its effects on the number of offspring produced by others who may have the same design—that individual's relatives—with each effect discounted by the appropriate measure of relatedness, often designated by *r* (Dawkins, 1982; Hamilton, 1964). Above, we used C_i and B_i to refer to effects on a design's Darwinian fitness; henceforth, we will use these variables to refer to effects on a design's inclusive fitness.[1]

Kin-directed helping behavior is common in the animal kingdom. But on occasion, one finds a species in which individuals help nonrelatives as well. How can a design feature that decreases one's own inclusive fitness while simultaneously increasing that of nonrelative be selected for? Although rare compared to kin-directed helping, such behavior does exist. For example, although kinship is a major predictor of food sharing in vampire bats, they share food with certain nonrelatives as well. Male baboons sometimes protect offspring not their own (Smuts, 1986). Unrelated chimpanzees will come to each other's aid when threatened (de Waal & Luttrell, 1988).

Williams (1966), Trivers (1971), Axelrod and Hamilton (1981), and Axelrod (1984) provided a second approach to the problem of altruism, reciprocal altruism theory, which in effect draws on the economist's concept of trade. Selection may act to create physiological or psychological mechanisms designed to deliver benefits even to nonrelatives, provided that the delivery of such benefits acts, with sufficient probability, to cause reciprocal benefits to be delivered in return. Such social exchange is easily understood as advantageous whenever there exist what economists call "gains in trade"—that is, whenever what each party receives is worth more than what it cost to deliver the reciprocal benefit to the other party. Ecologically realistic conditions, however, seldom provide opportunities in which two parties simultaneously have value to offer each other. For this reason, biologists have tended to focus on situations of deferred implicit exchange, where one party helps another at one point in time, in order to increase the probability that when their situations are reversed at some (usually) unspecified time in the future, the act will be reciprocated (hence the terms reciprocal altruism, reciprocation, or, as we prefer for the general class, social exchange).

If the reproductive benefit one receives in return is larger than the cost one incurred in rendering help, then individuals who engage in this kind of reciprocal helping behavior will outreproduce those who do not, causing this kind of helping design to spread. For example, if a vampire bat fails to find food for two nights in a row it will die, and there is high variance in food-gathering success. Sharing food allows the bats to cope with this variance, and the major predictor of whether a bat will share food with a nonrelative is whether the nonrelative has shared with that individual in the past (Wilkinson, 1988, 1990). Reciprocal altruism is simply cooperation between two or more individuals for mutual benefit, and it is variously known in the literature as social exchange, cooperation, or reciprocation. Design features that allow one to engage in reciprocal altruism can be selected for because they result in a net increase in one's own reproduction or that of one's relatives and, consequently, in the reproduction of the design features that produce this particular kind of cooperative behavior.

For example, according to reciprocal-altruism theory, cognitive programs that generate food sharing among nonrelatives can be selected for only if they exhibit certain design features. By cataloging these design features, Wilkinson (1988, 1990) was able to look for—and discover—heretofore unknown aspects of the psychology and behavior of female vampire bats. Reciprocal altruism theory guided his research program:

> I needed to demonstrate that five criteria were being met: that females associate for long periods, so that each one has a large but unpredictable number of opportunities to engage in blood sharing; that the likelihood of an individual regurgitating to a roostmate can be predicted on the basis of their past association; that the roles of donor and recipient frequently reverse; that the short-term benefits to the recipient are greater than the costs to the donor; and that donors are able to recognize and expel cheaters from the system. (Wilkinson 1990, p. 77)

Like kin selection theory, reciprocal altruism theory suggested a host of hypotheses about phenotypic design, which allowed animal behavior researchers to discover many previously unsuspected phenomena. Recently, it has done the same for those who study social exchange in humans. Reciprocal altruism theory has allowed researchers to derive a rich set of hypotheses about the design features of the cognitive programs that generate cooperative behavior in humans. We will examine some of these hypotheses and the evidence for them.

This chapter is divided into three parts. In the first part (Selection Pressures) we explore some of the constraints reciprocal altruism theory places on the class of designs that can evolve in humans. These "evolvability constraints" (see Tooby & Cosmides, this volume) led us to develop a set of hypotheses about the design features of the cognitive programs that are responsible for reasoning about social exchange. In the second part (Cognitive Processes) we review research that we and others have conducted to test these hypotheses and show that the cognitive programs that govern reasoning about social exchange in humans have many of the design features one would expect if they were adaptations sculpted by the selection pressures discussed in the first part. In the third part (Implications for Culture) we discuss the implications of this work for understanding cross-cultural uniformities and variability in cooperative behavior.

SELECTION PRESSURES

Natural selection permits the evolution of only certain strategies for engaging social exchange. To be selected for, a design governing reasoning about social exchange must embody one of these strategies—in other words, it must meet an "evolvability criterion" (see Tooby & Cosmides, this volume). By studying the nature of these strategies, one can deduce many properties that human algorithms regulating social exchange must have, as well as much about the associated capabilities such algorithms require to function properly. Using this framework, one can then make empirical predictions about human performance in areas that are the traditional concern of cognitive psychologists: attention, communication, reasoning, the organization of memory, and learning. One can also make specific predictions about human performance on reasoning tests, such as the ones we will discuss in Cognitive Processes (following).

In this part, we explore the nature of the selection pressures on social exchange during hominid evolution—the relevant evolvability constraints—and see what these allow one to infer about the psychological basis for social exchange in humans.

Game-Theoretic Constraints on the Evolution of Social Exchange

> The critical act in formulating computational theories turns out to be the discovery of valid constraints on the way the world is structured. (Marr & Nishihara, 1978, p. 41)

In *Evolution and the Theory of Games*, John Maynard Smith (1982) pointed out that natural selection has a game-theoretic structure. Alternative designs are selected for or not because of the different effects they have on their "own" reproduction—that is, on the reproduction of all identical designs in the population. Some designs will outreproduce others until they become universal in the population; others will be selected out. Using game theory, one can mathematically model this process with some precision. This is true whether one is describing the alternative designs anatomically, physiologically, or cognitively. For example, it is irrelevant to the analysis whether one describes a design change in a particular region of the brain anatomically—as an increase in the density of serotonin receptors in that region—physiologically—as an increase in the rate of serotonin uptake (which was caused by the increased receptor density)—or cognitively—as a difference in how the individual who has the increased receptor density processes information. All that matters to the analysis is what effect the design change—however described—has on its own reproduction. Because our concern in this chapter is the evolution of the information-processing mechanisms that generate cooperative behavior, we will describe alternative designs cognitively, by specifying the different rules that they embody and the representations that those rules act upon.

To see how a game-theoretic analysis works, consider how one can use it to understand the ramifications of reciprocal altruism theory for the evolution of social exchange between unrelated individuals.

Designs reproduce themselves through the reproduction of the individuals who embody them. Given an individual, *i*, define a *benefit to i* (B_i) as the extent to which any act, entity, or state of affairs increases the inclusive fitness of that individual. Similarly, define a *cost to i* (C_i) as the extent to which any act, entity, or state of affairs decreases the inclusive fitness of individual *i*. Let 0_i refer to any act, entity, or state of affairs that has no effect on *i*'s inclusive fitness. A cognitive program that causes a decrease in its own inclusive fitness while increasing that of an unrelated individual can evolve only if it has design features that embody the evolvability constraints of reciprocal altruism theory. A game-theoretic analysis allows one to explore what these constraints are. For ease of explication, the two interactants in a hypothetical social exchange will be designated "you" and "I," with appropriate possessive pronouns.

Reciprocal altruism, or social exchange, typically involves two acts: what "you" do for "me" (act 1), and what "I" do for "you" (act 2). For example, you might help me out by baby-sitting my child (act 1), and I might help you by taking care of your vegetable garden when you are out of town (act 2). Imagine the following situation: Baby-sitting my child inconveniences you a bit, but this inconvenience is more than compensated for by my watering your garden when you are out of town. Similarly, watering your garden inconveniences me a bit, but this is outweighed by the benefit to me of your baby-sitting my child. Formally put:

1. Your doing act 1 for me benefits me (B_{me}) at some cost to yourself (C_{you}).
2. My doing act 2 for you benefits you (B_{you}) at some cost to myself (C_{me}).
3. The benefit to you of receiving my act 2 is greater than the cost to you of doing act 1 for me (B_{you} of act 2 > C_{you} of act 1).
4. The benefit to me of receiving act 1 from you is greater than the cost to me of doing act 2 for you (B_{me} of act 1 > C_{me} of act 2).

If these four conditions are met—if acts 1 and 2 have this cost/benefit structure—then we would both get a net benefit by exchanging acts 1 and 2. Social exchange, or reciprocal altruism, is an interaction that has this mutually beneficial cost/benefit structure (see Table 3.1).

At first glance, one might think that natural selection would favor the emergence of cognitive programs with decision rules that cause organisms to participate in social exchange whenever the above conditions hold. After all, participation would result, by

Table 3.1 Sincere Social Contracts: Cost/Benefit Relations When One Party Is Sincere, and That Party Believes the Other Party Is Also Sincere[a]

	My offer: "If you do Act 1 for me then I'll do Act 2 for you."			
	Sincere offer		*Sincere acceptance*	
	I believe:		You believe:	
You do Act 1	B_{me}	C_{you}	B_{me}	C_{you}
You do not do Act 1	0_{me}	0_{you}	0_{me}	0_{you}
I do Act 2	C_{me}	B_{you}	C_{me}	B_{you}
I do not do Act 2	0_{me}	0_{you}	0_{me}	0_{you}
Profit margin	positive: $B_{me} > C_{me}$	positive: $B_{you} > C_{you}$	positive: $B_{me} > C_{me}$	positive: $B_{you} > C_{you}$
Translation of the offer into the value systems of the participants:				
My terms	"If B_{me} then C_{me}"		"If B_{me} then C_{me}"	
Your terms	"If C_{you} then B_{you}"		"If C_{you} then B_{you}"	

[a] B_x = benefit to *x*; C_x = cost to *x*; 0_x = no change in *x*'s zero-level utility baseline. The zero-level utility baseline is the individual's level of well-being (including expectations about the future) at the time the offer is made, but independent of it. Benefits and costs are increases and decreases in one's utility, relative to one's zero-level utility baseline.

definition, in a net increase in the replication of such designs, as compared with alternative designs that cause one to not participate.

But there is a hitch: You can benefit *even more* by cheating me. If I take care of your garden, but you do not baby-sit my child—i.e., if I cooperate by doing act 2 for you, but you defect on the agreement by not doing act 1 for me—then you benefit more than if we both cooperate. This is because your payoff for cheating when I have cooperated (B_{you}) is greater than your payoff for mutual cooperation ($B_{you} - C_{you}$)—you have benefited from my taking care of your garden without having inconvenienced yourself by baby-sitting for me. Moreover, the same set of incentives applies to me. This single fact constitutes a barrier to the evolution of social exchange, a problem that is structurally identical to one of the most famous situations in game theory: the one-move Prisoner's Dilemma (e.g., Axelrod, 1984; Axelrod & Hamilton, 1981; Boyd, 1988; Trivers, 1971).[2]

Mathematicians and economists use game theory to determine which decision rules will maximize an individual's monetary profits or subjective utility. Consequently, they express payoffs in dollars or "utils." Such currencies are inappropriate to an evolutionary analysis, however, because the goal of an evolutionary analysis is different. Evolutionary biologists use game theory to explore evolvability constraints. The goal is to determine which decision rules can, in principle, be selected for—which will, over generations, promote their own inclusive fitness. For this purpose, units of inclusive fitness are the only relevant payoff currency. Other assumptions are minimal. The organism need not "know," either consciously or unconsciously, *why* the decision rule it executes is better or worse than others, or even *that* the rule it executes is better or worse than others. To be selected for, a decision rule must promote its inclusive fitness better than alternative rules—and that's all. It doesn't need to make one happy, it doesn't need to maximize subjective utility, it doesn't need to promote the survival of the species, it doesn't need to promote social welfare. To be selected for, it need only promote its own replication better than alternative designs.

		you	
		C	D
me	C	me: R = +3 you: R = +3	me: S = −2 you: T = +5
	D	me: T = +5 you: S = −2	me: P = 0 you: P = 0

C = Cooperate
D = Defect
R = Reward for mutual cooperation
T = Temptation to defect
S = Sucker's payoff
P = Punishment for mutual defection

Constraints: T > R > P > S; R > (T+S)/2*

*For an interated game, R > (T + S)/2. This is to prevent players from "cooperating" to maximize their utility by alternately defecting on one another.

Figure 3.1 Payoff Schedule for the Prisoner's Dilemma situation in game theory.

Mathematicians and economists have used the Prisoner's Dilemma to understand how cooperation can arise in the absence of a "Leviathan," that is, a powerful state or agency that enforces contracts. Evolutionary biologists have used it to understand the conditions under which design features that allow individuals to cooperate can be selected for. It is a game in which mutual cooperation would benefit both players, but it is in the interest of each player, individually, to defect, cheat, or inform on the other. It is frequently conceptualized as a situation in which two people who have collaborated in committing a crime are prevented from communicating with each other, while a district attorney offers each individual a lighter sentence if he will snitch on his partner. But the payoffs can represent anything for which both players have a similar preference ranking: money, prestige, points in a game—even inclusive fitness. A possible payoff matrix and the relationship that must exist between variables is shown in Figure 3.1.

Looking at this payoff matrix, one might ask: "What's the dilemma? I will be better off, and so will you, if we both cooperate—you will surely recognize this and cooperate with me." If there is only one move in the game, however, *it is always in the interest* of each party to defect (Luce & Raiffa, 1957). That is what creates the dilemma, as we will show below.

Figure 3.2 shows that the cost/benefit structure of a social exchange creates the same payoff matrix as a Prisoner's Dilemma: $(B_i) > (B_i - C_i) > 0 > C_i$ (i.e., T > R > P > S); and $(B_i - C_i) > (B_i - C_i)/2$ (i.e., R > T + S/2). *In other words, if I* cooperate on our agreement, you get B_{you} for defecting, which is greater than the $B_{you} - C_{you}$ you would get for cooperating (i.e., T > R). If I defect on our agreement, you get nothing for defecting (this is equivalent to our not interacting at all; thus P = 0 and R > P), which is better than the C_{you} loss you would incur by cooperating (i.e., P > S). The payoffs are in inclusive fitness units—the numbers listed are included simply to reinforce the analogy with Figure 3.1. In actuality, there is no reason why C_{me} must equal C_{you} (or $B_{me} = B_{you}$); an exchange will have the stucture of a Prisoner's Dilemma as long as mutual cooperation would produce a net benefit for both of us.

Now that we have defined the situation, consider two alternative decision rules:

Decision rule 1: *Always cooperate.*
Decision rule 2: *Always defect.*

		you	
		C	D
me	C	me: $R = B_{me} - C_{me} = +3$ you: $R = B_{you} - C_{you} = +3$	me: $S = C_{me} = -2$ you: $T = B_{you} = +5$
	D	me: $T = B_{me} = +5$ you: $S = C_{you} = -2$	me: $P = 0_{me} = 0$ you: $P = 0_{you} = 0$

Figure 3.2 Social exchange sets up a Prisoner's Dilemma. B_i = Benefit to *i*, C_i = Cost to *i*, 0_i = *i*'s inclusive fitness is unchanged.

An individual with cognitive programs that embody decision rule 1 would be an indiscriminate cooperator; an individual with cognitive programs that embody decision rule 2 would be an indiscriminate cheater.

Now imagine a population of organisms, most of whom have cognitive programs that embody decision rule 1, but a few of whom have cognitive programs that embody decision rule 2.[3] Then imagine a tournament that pits the reproduction of decision rule 1 against that of decision rule 2.

In this tournament, both sets of individuals face similar environments. For example, one might specify that both types of organisms are subject to the same payoff matrix, that each organism participates in three interactions per "generation," and that these three interactions must be with three different individuals, randomly chosen from the population. After every organism has completed its three interactions, each organism "reproduces" and then "dies." "Offspring" carry the same decision rule as the "parent," and the number of offspring produced by an individual is proportional to the payoffs it gained in the three interactions it participated in in that generation. This process repeats itself every generation.

Using this tournament, one can ask, After one generation, how many replicas of rule 1 versus rule 2 exist in the population? How many replicas of each rule exist after *n* generations? If one were to run a computer model of this tournament, one would find that after a few generations individuals who operate according to rule 2 ("Always defect") would, on average, be leaving more offspring than individuals operating according to rule 1 ("Always cooperate"); the magnitude of the difference between them is rule 2's "selective advantage" over rule 1. This magnitude will depend on what payoff and opportunity parameters were specified in the program used, as well as the population composition.

After a larger number of "generations," rule 1—"Always cooperate"—would be selected out. For every interaction with a cheater, rule 1 would lose two inclusive fitness points, and rule 2 would gain five. Consequently, indiscriminate cooperators would eventually be selected out, and indiscriminate cheaters would spread through the population; the number of generations this would take is a function of how many cheaters versus indiscriminate cooperators were in the initial population. In practice, a population of "cheaters" is a population of individuals who never participate in social exchange; if you "cheat" by not doing act 1 for me, and I "cheat" by not doing act 2 for you, then, in effect, we have exchanged nothing. And an indiscriminate cooperator in the midst of defectors is, in practice, always an "altruist" or victim, continually incurring costs in the course of helping others, but receiving no benefits in return.

So, after *n* generations, where *n* is a function of the magnitude of rule 2's selective advantage in the tournament's "environment" and other population parameters, one would find that rule 2 had "gone to fixation": Virtually all individuals would have rule 2, and, regardless of the population's absolute size, a vanishingly small proportion of the individuals in it would have rule 1.[4]

By using this kind of logic, one can show that if a new design coding for rule 2—"Always defect"—were to appear in a population that is dominated by individuals with rule 1—"Always cooperate"—it would spread through the population until it became fixed, and it would not be vulnerable to invasion by rule 1 (see, e.g., Axelrod, 1984). In a tournament pitting indiscriminate altruists against indiscriminate cheaters, the cheaters will come to dominate the population.

One might object that real life is not like a Prisoner's Dilemma, because real-life exchanges are simultaneous, face-to-face interactions. You can directly recognize whether I am about to cheat you or not (provided you are equipped with cognitive equipment that guides you into making this discrimination). If I show up without the item I promised, then you simply do not give me what I want. This is often true in a twentieth-century market economy, where money is used as a medium of exchange. But no species that engages in social exchange, including our own, evolved the information-processing mechanisms that enable this behavior in the context of a market economy with a medium of exchange.

Virtually any nonsimultaneous exchange increases the opportunity for defection, and in nature, most opportunities for exchange are not simultaneous. For example, a drowning man needs immediate assistance, but while he is being pulled from the water, he is in no position to help his benefactor. Opportunities for simultaneous mutual aid—and therefore for the withdrawal of benefits in the face of cheating—are rare in nature for several reasons:

The "items" of exchange are frequently acts that, once done, cannot be undone (e.g., protection from an attack and alerting others to the presence of a food source).

The needs and abilities of organisms are rarely exactly and simultaneously complementary. For example, a female baboon is not fertile when her infant needs protection, yet this is when the male's ability to protect is of most value to her.

On those occasions when repayment is made in the same currency, simultaneous exchange is senseless. If two hunters both make kills on the same day, they gain nothing from sharing their kills with each other: They would be swapping identical goods. In contrast, repayment in the same currency can be advantageous when exchange is not simultaneous, because of declining marginal utilities: The value of a piece of meat is larger to a hungry individual than to a sated one.

Thus, in the absence of a widely accepted medium of exchange, most exchanges are not simultaneous and therefore do provide opportunities for defection. You must decide whether to benefit me or not without any guarantee that I will return the favor in the future. This is why Trivers (1971) describes social exchange in nature as "reciprocal altruism." I behave "altruistically" (i.e., I incur a cost in order to benefit you) at one point in time, on the possibility that you will reciprocate may altruistic act in the future. If you do, in fact, reciprocate, then our "reciprocally altruistic" interaction is properly described as an instance of delayed mutual benefit: Neither of us has incurred a net cost; both of us have gained a net benefit.

A system of mutual cooperation cannot emerge in a one-move Prisoner's Dilemma because it is always in the interest of each player to defect. In fact, the argument is general to any known, fixed number of games (Luce & Raiffa, 1957). But selection pressures change radically when individuals play a *series* of Prisoner's Dilemma games. Mutual cooperation—and therefore social exchange—can emerge between two players when (a) there is a high probability that they will meet again, (b) neither knows for sure exactly how many times they will meet,[5] and (c) they do not value later payoffs by too much less than earlier payoffs (Axelrod, 1984; Axelrod & Hamilton, 1981). If the parties are making a series of moves rather than just one, then one party's behavior on a move can influence the other's behavior on future moves. If I defect when you cooperated, then you can retaliate by defecting on the next move; if I cooperate, then you can reward me by cooperating on the next move. In an iterated Prisoner's Dilemma game, a system can emerge that has incentives for cooperation and disincentives for defection.

The work of Trivers (1971), Axelrod and Hamilton (1981), and Axelrod (1984) has shown that indiscriminate cooperation (Decision rule 1) cannot be selected for when the opportunity for cheating exists. But *selective* cooperation can be selected for. Decision rules that cause one to cooperate with other cooperators and defect on cheaters can invade a population of noncooperators.

Consider, for example, Decision rule 3:

> Decision rule 3: *Cooperate on the first move; on subsequent moves, do whatever your partner did on the previous move.*

This decision rule is known in the literature as TIT FOR TAT (Axelrod & Hamilton, 1981). If rule 3's partner cooperates on a move, rule 3 will cooperate on the next move with that partner. If rule 3's partner defects on a move, rule 3 will defect on the next move with that partner. It has been shown that rule 3 can invade a population dominated by indiscriminate cheaters (individuals who behave according to decision rule 2: "Always defect"). Using the payoff matrix in Figure 3.2, it is clear that rule 3 would outreproduce rule 2: Mutual cooperators (pairs of individuals who behave according to rule 3) would get strings of +3 inclusive fitness points, peppered with a few −2s from a first trial with a rule 2 cheater (after which the cooperator would cease to cooperate with that individual). In contrast, mutual defectors (pairs of individuals who behave according to rule 2) would get strings of 0s, peppered with a few +5s from an occasional first trial with a rule 3 cooperator (after which the cooperator would never cooperate with that individual again).

Game-theoretic analyses have shown that a decision rule embodying a cooperative strategy can invade a population of noncooperators if, and only if, it cooperates with other cooperators and excludes (or retaliates against) cheaters. If a decision rule regulating when one should cooperate and when one should cheat violates this constraint, then it will be selected against.

Axelrod (1984) has shown that there are many decision rules that do embody this constraint. All else equal (an important caveat), any of these could, in theory, have been selected for in humans. Which decision rule, out of this constrained set, is embodied in the cognitive programs that actually evolved in the human lineage is an empirical question. But note that to embody any of this class of decision rules, the cognitive programs involved would have to incorporate a number of specific design features:

1. They must include algorithms that are sensitive to cues that indicate when an exchange is being offered and when reciprocation is expected.
2. They must include algorithms that estimate the costs and benefits of various actions, entities, or states of affairs to oneself.[6]
3. They must include algorithms that estimate the costs and benefits of various actions, entities, or states of affairs to others (in order to know when to initiate an exchange).
4. They must include algorithms that estimate the probability that these actions, entities, or states of affairs will come about in the absence of an exchange.
5. They must include algorithms that compare these estimates to one another (in order to determine whether $B_i > C_i$).
6. They must include decision rules that cause *i* to reject an exchange offer when $B_i < C_i$.
7. They must include decision rules that cause *i* to accept (or initiate) an exchange when $B_i > C_i$ (and other conditions are met).
8. They must include algorithms with inference procedures that capture the intercontingent nature of exchange (see Cosmides & Tooby, 1989, pp. 81–84).
9. They must include algorithms that can translate the exchange into the value assignments appropriate to each participant.
10. They must include algorithms that can detect cheaters (these must define cheating as an illicitly taken benefit).
11. They must include algorithms that cause one to punish cheating under the appropriate circumstances.
12. They must include algorithms that store information about the history of one's past exchanges with other individuals (in order to know when to cooperate, when to defect, and when to punish defection).
13. They must include algorithms that can recognize different individuals (in order to do any of the above).
14. They need not include algorithms for detecting indiscriminate altruists, because there shouldn't be any.

Not all of these algorithms need to be part of the same "mental organ." For example, because algorithms that can do 2, 3, 4, 5, and 13 are necessary to engage in social interactions other than exchange—such as aggressive threat—these might be activated even when the algorithms that are specific to social exchange are not.

Design features 1 to 14 are just a partial listing, based on some very general constraints on the evolution of social exchange that fall out of an examination of the iterated Prisoner's Dilemma. These constraints are general in the sense that they apply to the evolution of reciprocal altruism in almost any species—from reciprocal egg trading in hermaphroditic fish (Fischer, 1988) to food sharing in humans. Other, species-specific constraints on the design of social exchange algorithms can be derived by considering how these general selection pressures would have manifested themselves in the ecological context of hominid evolution.

For example, the sharing rules that are applied to high-variance resources, such as hunted meat, should differ in some ways from those that are applied to low-variance resources, such as gathered plant foods (Kaplan & Hill, 1985; see also Implications for Culture, this chapter). This raises the possibility that human exchange algorithms have two alternative, context-specific modes of activation. Both modes would have to satisfy the general constraints listed above, but they might differ considerably in various details, such as whether one expects to be repaid in the same currency (e.g., meat for meat), whether one requires reciprocation before one is willing to help a second time, or whether one is quick to punish suspected cheaters. Another example of how ecological context can place species-specific constraints on design concerns the kind of *representations that exchange algorithms can be expected to operate on.* For example, exchange algorithms in humans should operate on more abstract representations than exchange algorithms in vampire bats. The reciprocation algorithms of vampire bats could, in principle, operate on representations of regurgitated blood, because this is the only item that they exchange. But item-specific representations of this kind would not make sense for the exchange algorithms of humans. Because our ancestors evolved the ability to make and use tools and to communicate information verbally, exchange algorithms that could accept a wide and ever-changing variety of goods, services, and information as input would enjoy a selective advantage over ones that were limited to only a few items of exchange. To accommodate an almost limitless variety of inputs—stone axes, meat, help in fights, sexual access, information about one's enemies, access to one's water hole, necklaces, blow guns, and so forth—representations of particular items of exchange would have to be translated into an abstract "lingua franca" that the various exchange algorithms could operate on. This constraint led us to hypothesize that an item-specific repesentation of an exchange would be translated into more abstract cost-benefit representations (like those in the last two lines of Table 3.1) at a relatively early stage in processing, and that many of the algorithms listed earlier would operate on these cost-benefit representations (Cosmides & Tooby, 1989). Because some of these species-specific constraints on the evolution of social exchange in humans have interesting implications for cultural variation, we will defer a discussion of them to the third part of this chapter (Implications for Culture), where we discuss social exchange and culture.

David Marr argued that "an algorithm is likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is embodied" (1982, p. 27). This is because the nature of the problem places constraints on the class of designs capable of solving it. The iterated Prisoner's Dilemma is an abstract description of the problem of altruism between nonrelatives. By studying it, one can derive a set of general constraints that the cognitive problems of virtually any species must satisfy to be selected for under these circumstances. By studying the ecological context in which this problem manifested itself for our Pleistocene ancestors, one can derive additional constraints. All these constraints on the evolution of social exchange—those that apply across species and those that apply just to humans—allow one to develop a task analysis or, to use Marr's term, a "computational theory" of the adaptive problem of social exchange. Cosmides and Tooby (1989) used some of these constraints to develop the beginnings of a computational theory of social exchange, which we call "social contract theory." So as not to repeat ourselves here, we refer the reader to that article for details. By constraining the class of possible designs, this theory allowed us and others to make some predictions about the design features of the algorithms and representations that evolved to solve the problem of social exchange in humans. Design features 1–14 listed earlier are a small subset of those predictions.

The computational theory we developed has guided our research program on human reasoning. We have been conducting experiments to see whether people have cognitive processes that are specialized for reasoning about social exchange. The experiments we will review in the following part were designed to test for design features 1, 9, 10, and 14, as well as some other predictions derived from the computational theory. We have been particularly interested in testing the hypothesis that humans have algorithms that are specialized for detecting cheaters in situations of social exchange.

COGNITIVE PROCESSES

Differential reproduction of alternative designs is the engine that drives natural selection: If having a particular mental structure, such as a rule of inference, allows a design to outreproduce other designs that exist in the species, then that mental structure will be selected for. Over many generations it will spread through the population until it becomes a universal, species-typical trait.

Traditionally, cognitive psychologists have assumed that the human mind includes only general-purpose rules of reasoning and that these rules are few in number and content-free. But a cognitive perspective that is informed by evolutionary biology casts doubt on these assumptions. This is because natural selection is also likely to have produced many mental rules that are specialized for reasoning about various evolutionarily important domains, such as cooperation, aggressive threat, parenting, disease avoidance, predator avoidance, object permanence, and object movement. Different adaptive problems frequently have different optimal solutions, and can therefore be solved more efficiently by the application of different problem-solving procedures. When two adaptive problems have different optimal solutions, a single general solution will be inferior to two specialized solutions. In such cases, a jack-of-all-trades will necessarily be a master of none, because generality can be achieved only by sacrificing efficiency. Indeed, it is usually more than efficiency that is lost by being limited to a general-purpose method—generality may often sacrifice the very possibility of successfully solving a problem, as, for example, when the solution requires supplemental information that cannot be sensorily derived (this is known as the "frame problem" in artificial intelligence research).

The same principle applies to adaptive problems that require reasoning: There are cases where the rules for reasoning adaptively about one domain will lead one into serious error if applied to a different domain. Such problems cannot, in principle, be solved by a single general-purpose reasoning procedure. They are best solved by different special-purpose reasoning procedures.

For example, the rules of inference of the propositional calculus (formal logic) are general-purpose rules of inference: They can be applied regardless of what subject matter one is reasoning about. Yet the consistent application of these rules of logical reasoning will not allow one to detect cheaters in situations of social exchange, because what counts as cheating does not map onto the definition of violation imposed by the propositional calculus. Suppose you and I agree to the following exchange: "If you give me your watch then I'll give you $20." You would have violated our agreement—you would have cheated me—if you had taken my $20 but not given me your watch. But according to the rules of inference of the propositional calculus, the only way this rule can be violated is by your giving me your watch but my not giving you $20.[7] If the only mental rules my mind contained were the rules of inference of the propositional calculus, then I would not be able to tell when you had cheated me. Similarly, rules of inference for detecting cheaters on social contracts will not allow one to detect bluffs or double crosses in situations of aggressive threat (Cosmides & Tooby, in prep., a). What counts as a violation differs for a social contract, a threat, a rule describing the state of the world, and so on. Because of this difference, the same reasoning procedure cannot be successfully applied to all of these situations. As a result, there cannot be a general-purpose reasoning procedure that works for all of them. If these problems are to be solved at all, they must be solved by different specialized reasoning procedures.

Given the selection pressures discussed earlier, we can define a social contract as a situation in which an individual is obligated to satisfy a requirement of some kind, usually at some cost to him- or herself, in order to be entitled to receive a benefit from another individual (or group). The requirement is imposed because its satisfaction creates a situation that benefits the party that imposed it. Thus, a well-formed social contract expresses an intercontingent situation of mutual benefit: To receive a benefit, an individual (or group) is required to provide a benefit. Usually (but not always) one incurs a cost by satisfying the requirement. But that cost is outweighed by the benefit one receives in return.

Cheating is a violation of a social contract. A cheater is an individual who illicitly benefits himself or herself by taking a benefit without having satisfied the requirement that the other party to the contract made the provision of that benefit contingent on. In this section, we review evidence that people have cognitive adaptations that are specialized for reasoning about social contracts. We will pay particular attention to the hypothesis that people have inference procedures specialized for cheater detection.

Adaptations are aspects of the phenotype that were designed by natural selection. To show that an aspect of the phenotype is an adaptation, one must produce evidence that it is well designed for solving an adaptive problem. Contrary to popular belief, developmental evidence is not criterial: Adaptations need not be present from birth (e.g., breasts), they need not develop in the absence of learning or experience (e.g., vision, language—see Pinker & Bloom, this volume),[8] and they need not be heritable (Tooby & Cosmides, 1990a). In fact, although the developmental processes that create adaptations are inherited, adaptations will usually exhibit low heritability. Differences between individuals will not be due to differences in their genes because adaptations are, in most cases, universal and species-typical—everyone has the genes that guide their development. The filter of natural selection does not sift designs on the basis of their developmental trajectory per se:[9] It doesn't matter how a design was built, only *that* it was built, and to the proper specifications.

To say that an organism has cognitive procedures that are adaptations for detecting cheaters, one must show that these procedures are well designed for detecting cheaters on social contracts. One must also show that their design features are not more parsimoniously explained as by-products of cognitive processes that evolved to solve some other kind of problem, or a more general class of problems. We approached this question by studying human reasoning. A large literature already existed that showed that people are not very good at detecting violations of conditional rules, even when

these rules deal with familiar content drawn from everyday life. To show that people who ordinarily cannot detect violations of conditional rules can do so when that violation represents cheating on a social contract would constitute evidence that people have reasoning procedures that are specially designed for detecting cheaters in situations of social exchange.

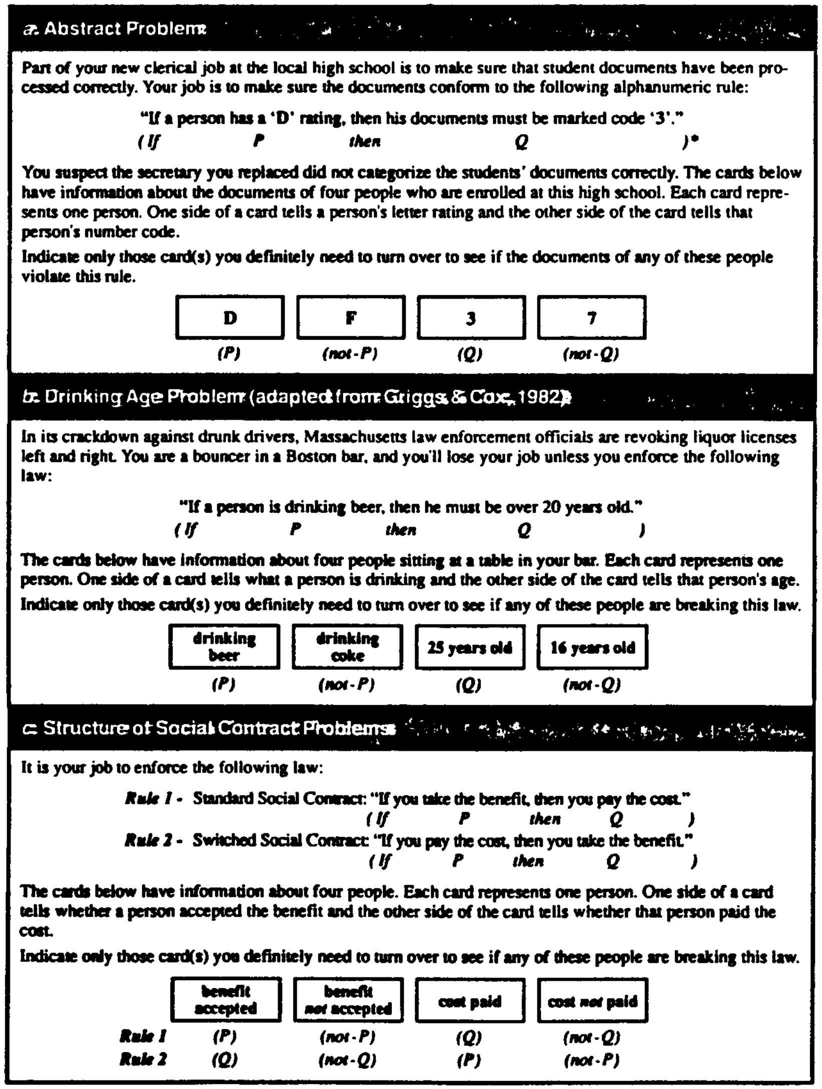
The Wason Selection Task

One of the most intriguing and widely used experimental paradigms for exploring people's ability to detect violations of conditional rules has been the Wason selection task (Wason, 1966; see Figure 3.3, panel a). Peter Wason was interested in Karl Popper's view that the structure of science was hypothetico-deductive. He wondered if everyday learning was really hypothesis testing—i.e., the search for evidence that contradicts a hypothesis. Wason devised his selection task because he wanted to see whether people are well equipped to test hypotheses by looking for evidence that could potentially falsify them. In the Wason selection task, a subject is asked to see whether a conditional hypothesis of the form *If P then Q* has been violated by any one of four instances represented by cards.

A hypothesis of the form *If P then Q* is violated only when *P* is true but *Q* is false: The rule in Figure 3.3, panel a, for example, can be violated only by a card that has a D on one side and a number other than 3 on the other side. Thus, one would have to turn over the *P* card (to see if it has a *not-Q* on the back) and the *not-Q* card (to see if it has a *P* on the back)—D and 7, respectively, for the rule in Figure 3.3, panel a. Consequently, the logically correct response for a rule of the form *If P then Q* is always *P & not-Q*.

Wason expected that people would be good at detecting violations of conditional rules. Nevertheless, over the past 25 years, he and many other psychologists have found that few people actually give this logically correct answer (less than 25% for rules expressing unfamiliar relations). Most people choose either the *P* card alone or *P & Q*. Few people choose the *not-Q* card, even though a *P* on the other side of it would falsify the rule.

A wide variety of conditional rules that describe some aspect of the world ("descriptive rules") have been tested; some of these have expressed relatively familiar relations, such as "If a person goes to Boston, then he takes the subway" or "If a person eats hot chili peppers, then he will drink a cold beer." Others have expressed unfamiliar relations, such as "If you eat duiker meat, then you have found an ostrich eggshell" or "If there is an 'A' on one side of a card, then there is a '3' on the other side." In many experiments, performance on familiar descriptive rules is just as low as on unfamiliar ones. For example, rules relating food to drink, such as the hot chili pepper rule above, have never elicited logical performance higher than that elicited by unfamiliar rules, even though the typical sophomore in such experiments has had about 22,000 experiences in which he or she has had both food and drink, and even though recurrent relations between certain foods and certain drinks are common—cereal with orange juice at breakfast, red wine with red meat, coffee with dessert, and so on. Sometimes familiar rules do elicit a higher percentage of logically correct responses than unfamiliar ones, but even when they do, they typically elicit the logically correct response from fewer than half of the people tested. For example, in the Wason selection task literature, the transportation problem—"If a person goes to Boston, then he takes the



a. Abstract Problem

Part of your new clerical job at the local high school is to make sure that student documents have been processed correctly. Your job is to make sure the documents conform to the following alphanumeric rule:

"If a person has a 'D' rating, then his documents must be marked code '3'."
*(If P then Q)**

You suspect the secretary you replaced did not categorize the students' documents correctly. The cards below have information about the documents of four people who are enrolled at this high school. Each card represents one person. One side of a card tells a person's letter rating and the other side of the card tells that person's number code.

Indicate only those card(s) you definitely need to turn over to see if the documents of any of these people violate this rule.

D	F	3	7
(P)	*(not-P)*	*(Q)*	*(not-Q)*

b. Drinking Age Problem (adapted from Griggs & Cox, 1982)

In its crackdown against drunk drivers, Massachusetts law enforcement officials are revoking liquor licenses left and right. You are a bouncer in a Boston bar, and you'll lose your job unless you enforce the following law:

"If a person is drinking beer, then he must be over 20 years old."
(If P then Q)

The cards below have information about four people sitting at a table in your bar. Each card represents one person. One side of a card tells what a person is drinking and the other side of the card tells that person's age.

Indicate only those card(s) you definitely need to turn over to see if any of these people are breaking this law.

drinking beer	drinking coke	25 years old	16 years old
(P)	*(not-P)*	*(Q)*	*(not-Q)*

c. Structure of Social Contract Problems

It is your job to enforce the following law:

Rule 1 - Standard Social Contract: "If you take the benefit, then you pay the cost."
(If P then Q)
Rule 2 - Switched Social Contract: "If you pay the cost, then you take the benefit."
(If P then Q)

The cards below have information about four people. Each card represents one person. One side of a card tells whether a person accepted the benefit and the other side of the card tells whether that person paid the cost.

Indicate only those card(s) you definitely need to turn over to see if any of these people are breaking this law.

	benefit accepted	benefit *not* accepted	cost paid	cost *not* paid
Rule 1	*(P)*	*(not-P)*	*(Q)*	*(not-Q)*
Rule 2	*(Q)*	*(not-Q)*	*(P)*	*(not-P)*

*The logical categories (*P*s and *Q*s) marked on the rules and cards are here only for the reader's benefit; they never appear on problems given to subjects.

Figure 3.3 Content effects on the Wason selection task. The logical structures of these three Wason selection tasks are identical: They differ only in propositional content. Regardless of content, the logical solution to all three problems is the same: *P* & *not-Q*. Although <25% of college students choose both these cards for the abstract problem in panel a, about 75% do for the drinking age problem (panel b)—a familiar "standard" social contract. Panel c shows the abstract structure of a social contract problem. Cheater detection algorithms should cause subjects to choose the "benefit accepted" card and the "cost not paid" card, *regardless of which logical categories they represent.* These cards correspond to the logical categories *P* & *not-Q* for the "standard" form in rule 1, but to the logical categories *not-P* & *Q* for the "switched" form in rule 2.

subway"—is the familiar descriptive rule that has the best record for eliciting logically correct responses, and performance on this rule was consistently higher in Cosmides's (1989) experiments than in most others. Even so, it elicited the logically correct *P & not-Q* response from only 48% of the 96 subjects tested. Recently, Stone and Cosmides (in prep.) tested rules expressing causal relations; the pattern of results is essentially the same as for descriptive rules. Humans do not appear to be naturally equipped to seek out violations of descriptive or causal rules.

When subjects are asked to look for violations of conditional rules that express social contracts, however, their performance changes radically. Consider the "drinking age problem" in Figure 3.3, panel b. It expresses a social contract in which one is entitled to receive a benefit (beer) only if one has satisfied a requirement (being a certain age).[10] The drinking age problem elicits the logically correct response, *P & not-Q*, from about 75% of subjects (e.g., Griggs & Cox, 1982; Cosmides, 1985). On this rule, it is very easy to see why one needs to check what the 16-year-old is drinking (the *not-Q* card) and the age of the person drinking beer (*P*), and it is equally obvious why the 25-year-old (*Q*) and the coke drinker (*not-P*) need not be checked. Experiments with the drinking age problem and other familiar social contracts show that human reasoning changes dramatically depending on the subject matter one is reasoning about. Such changes are known as "content effects."

Figure 3.3, panel c, shows the abstract structure of a social contract problem. To detect cheaters on a social contract, one would always want to choose the "benefit accepted" card and the "cost not paid" card, regardless of what logical category these cards happen to correspond to. For the drinking age problem, these cards happen to correspond to the logical categories *P* and *not-Q*. Consequently, a subject who was looking for cheaters on a social contract would, by coincidence, choose the logically correct answer. But, as this figure also shows, there are situations in which the correct answer if one is looking for cheaters is not the logically correct answer. This point is important, and we will return to it later.

When we began this research in 1983, the literature on the Wason selection task was full of reports of a wide variety of content effects, and there was no satisfying theory or empirical generalization that could account for these content effects. When we categorized these content effects according to whether they conformed to social contracts, a striking pattern emerged. Robust and replicable content effects were found only for rules that related terms that are recognizable as benefits and cost/requirements in the format of a standard social contract—i.e., in the format of Rule 1 in Figure 3.3, panel c (see Cosmides, 1985, for a review). No thematic rule that was not a social contract had ever produced a content effect that was both robust and replicable. Moreover, most of the content effects reported for non-social contract rules were either weak, clouded by procedural difficulties, or had some earmarks of a social contract problem. All told, for non-social contract thematic problems, 3 experiments had produced a substantial content effect, 2 had produced a weak content effect, and 14 had produced no content effect at all. The few effects that were found did not replicate. In contrast, 16 out of 16 experiments that fit the criteria for standard social contracts—i.e., 100%—elicited substantial content effects. (Since that time, additional types of content have been tested; evolutionarily salient contents have elicited new, highly patterned content effects that are indicative of additional cognitive specializations in reasoning.)[11]

Special design is the hallmark of adaptation. As promising as these initial results were, they were not sufficient to demonstrate the existence of an adaptation for social exchange. First, although the familiar rules that were social contracts always elicited a robust effect, and the familiar rules that were not social contracts failed to elicit a robust and replicable effect, this was true *across* experiments; individual experiments usually pitted performance on a familiar social contract against performance on an unfamiliar descriptive rule. Because they confounded familiarity with whether a rule was a social contract or not, these experiments could not decide the issue.[12] Familiarity could still be causing the differences in performance in complex ways that varied across different subject populations. Second, even if it were shown that familiarity could not account for the result, these experiments could not rule out the hypothesis that social contract content simply facilitates *logical* reasoning. This is because the adaptively correct answer, if one is looking for cheaters on the social contracts tested, happened to also be the logically correct answer. To show that the results were caused by rules of inference that are specialized for reasoning about social exchange, one would have to test social contract rules in which the correct "look for cheaters" answer is *different* from the logically correct answer.

Below, we review evidence addressing these, and other, hypotheses. Our goal will be twofold: (a) to show that the reasoning procedures involved show the features of special design that one would expect if they were adaptations for social exchange, and (b) to show that the results cannot be explained as by-products of other, more general-purpose reasoning procedures.

Did the Social Contract Problems Elicit Superior Performance Because They Were Familiar?

Familiar social contracts elicited high levels of apparently logical performance. Could this result be a by-product of the familiarity of the social contract rules tested? Suppose we have general-purpose reasoning procedures whose design makes us more likely to produce logically correct answers for familiar or thematic rules. Then high levels of *P & not-Q* responses to familiar social contract rules could be a by-product of the operation of these general-purpose mechanisms, rather than the result of algorithms specialized for reasoning about social exchange.

The first family of hypotheses that we tested against was the availability theories of reasoning, which are sophisticated and detailed versions of this "by-product" hypothesis (Griggs & Cox, 1982; Johnson-Laird, 1982; Manktelow & Evans, 1979; Pollard, 1982; Wason, 1983; for a detailed review, see Cosmides, 1985). These theories come in a variety of forms with some important theoretical differences, but common to all is the notion that the subject's actual past experiences create associational links between terms mentioned in the selection task. These theories sought to explain the "now you see it, now you don't" results common for certain familiar descriptive rules, such as the transportation problem. Sometimes these rules elicited a small content effect; other times they elicited none at all. This was in spite of the fact that the *relations* tested—between, for example, destinations and means of transportation or between eating certain foods in conjunction with certain drinks—were perfectly familiar to subjects. This meant that a general familiarity with the relation itself was not sufficient to explain the results. The proposal, therefore, was that subjects who had, for example, gone to Boston more often by cab than by subway would be more likely to pick "Boston" and "cab"—i.e., *P & not-Q*—for the rule "If one goes to Boston, then one takes the subway" than those who had gone to Boston more often by subway (*Q*). According to the availability theories, the more exposures a subject has had to the co-occurrence of *P* and *Q*, the stronger that association will become, and the easier it will come to mind, i.e., become "available" as a response. A subject is more likely to have actually experienced the co-occurrence of *P* and *not-Q* for a familiar rule, therefore familiar rules are more likely to elicit logically correct responses than unfamiliar rules. But whether a given rule elicits a content effect or not will depend on the actual, concrete experiences of the subject population tested.

Despite their differences, the various availability theories make the same prediction about unfamiliar rules. If all the terms in a task are unfamiliar, the only associational link available will be that created between *P* and *Q* by the conditional rule itself, because no previous link will exist among any of the terms. Thus *P & Q* will be the most common response for unfamiliar rules. *P & not-Q* responses will be rare for all unfamiliar rules, whether they are social contracts or not. The fact that a social-contract-type *relation* might be familiar to subjects is irrelevant: Previous results had already shown that the familiarity of a relation could not, by itself, enhance performance.

We can test against this family of hypotheses because social contract theory makes very different predictions. If people do have inference procedures that are specialized for reasoning about social contracts, then these ought to function, in part, as frame or schema builders, which structure new experiences. This means they should operate in *unfamiliar* situations. No matter how unfamiliar the relation or terms of a rule, if the subject perceives the terms as representing a rationed benefit and a cost/requirement in the implicational arrangement appropriate to a social contract, then a cheater detection procedure should be activated. Social contract algorithms need to be able to operate in new contexts if one is to be able to take advantage of new exchange opportunities. Therefore, the ability to operate on social contracts even when they are unfamiliar is a design feature that algorithms specialized for reasoning about social exchange should have. Social contract theory predicts a high level of *P & not-Q* responses on all "standard" social contract problems, whether they are familiar or not. It is silent on whether availability exerts an independent effect on non-social contract problems. A standard social contract is one that has the abstract form, *If you take the benefit, then you pay the cost* (see Rule 1, Figure 3.3, panel c).

In the first set of experiments, we pitted social contract theory against the availability family of theories by testing performance on an *unfamiliar* standard social contract—a problem for which the two hypotheses make diametrically opposite predictions (for details, see Cosmides, 1989, Experiments 1 and 2). Each subject was given four Wason selection tasks to solve: an unfamiliar social contract, an unfamiliar descriptive rule, a familiar descriptive rule (the transportation problem), and an abstract problem (as in Figure 3.3, panel a). Problem order was counterbalanced across subjects. The abstract problem was included because it is the usual standard for assessing the presence of a content effect in the Wason selection task literature; the transportation problem was included as a standard against which to judge the size of any social contract effect that might occur.

Rules such as, "If you eat duiker meat, then you have found an ostrich eggshell," or, "If a man eats cassava root, then he must have a tattoo on his face," were used for the unfamiliar problems; we felt it was safe to assume that our subjects would not have associative links between terms such as "cassava root" and "no tattoo" stored in long-term memory. An unfamiliar rule was made into a social contract or a descriptive rule by manipulating the surrounding story context. For example, a social contract version of the cassava root rule might say that in this (fictitious) culture, cassava root (*P*) is a much prized aphrodisiac whereas molo nuts (*not-P*) are considered nasty and bitter, thereby conveying that eating cassava root is considered a benefit compared to eating molo nuts, which are the alternative food. Having a tattoo on one's face (*Q*) means one is married; not having a tattoo (*not-Q*) means one is unmarried. As subjects know that marriage is a contract in which certain obligations are incurred to secure certain benefits (many of which involve sexual access), being married in this story is the cost/requirement. Finally, the story explains that because the people of this culture are concerned about sexual mores, they have created the rule, "If a man eats cassava root, then he must have a tattoo on his face." The four cards, each representing one man, would read "eats cassava root," "eats molo nuts," "tattoo," and "no tattoo." Other story contexts were invented for other rules.

The descriptive version of the unfamiliar rule would also give meaning to the terms and suggest a meaningful relation between them, but the surrounding story would not give the rule the cost/benefit structure of a social contract. For example, it might explain that cassava root and molo nuts are both staple foods eaten by the people of the fictitious culture (i.e., there is no differential benefit to eating one over the other), but that cassava root grows only at one end of the island they live on, whereas molo nuts grow only at the other end. Having a tattoo on your face or not again indicates whether a man is married, and it so happens that married men live on the side of the island where the cassava root grows, whereas unmarried men live on the side where the molo nuts grow. Note that in this version, being married has no significance as a cost/requirement; it is merely a correlate of where one lives. The story then provides a meaningful relation to link the terms of the rule, "If a man eats cassava root, then he must have a tattoo on his face," by suggesting that it simply describes the fact that men are eating the foods that are most available to them.

Subjects who were given a cassava root version of the social contract rule were given a duiker meat version of the descriptive rule, and vice versa, so that no subject encountered two versions of the exact same unfamiliar rule. The availability theories predict low levels of *P & not-Q* responses on both unfamiliar rules, whether they are portrayed as social contracts or not. Social contract theory predicts high levels of *P & not-Q* responses for the unfamiliar social contract, but not for the unfamiliar descriptive rule. The predictions of the two theories, and the results of two different sets of experiments, are shown in Figure 3.4.

The results clearly favor social contract theory. Even though they were unfamiliar and culturally alien, the social contract problems elicited a high percentage of *P & not-Q* responses. In fact, both we and Gigerenzer and Hug (in press) found that the performance level for unfamiliar social contracts is just as high as it usually is for familiar social contracts such as the drinking age problem—around 75% correct in our experiments. Unfamiliar social contracts elicited levels of *P & not-Q* responses that were even higher than those elicited by the familiar descriptive transporation problem. From our various experiments, we estimated the size of the social contract effect to be about 1.49 times larger than the size of the effect that availability has on familiar descriptive problems.

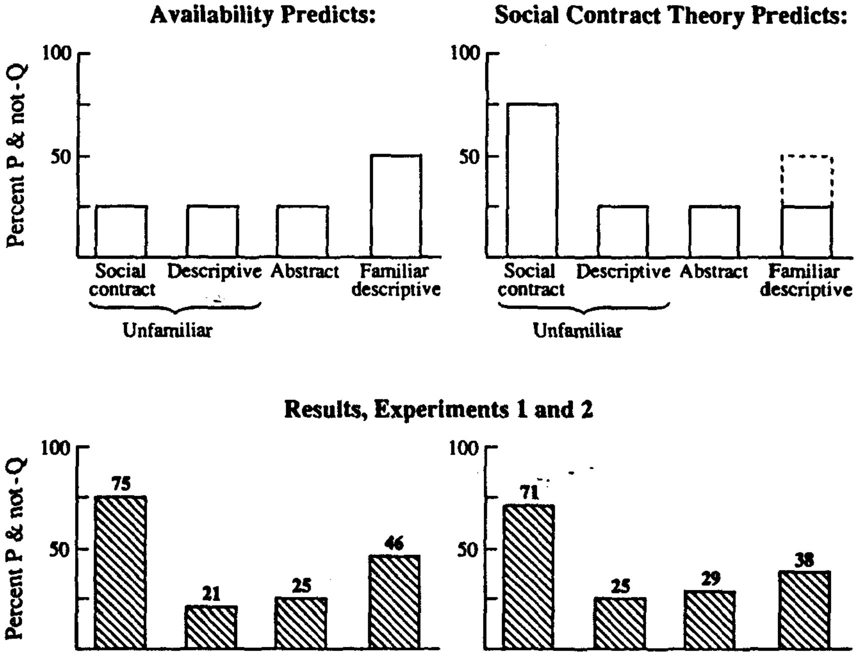


Figure 3.4 Social contract theory versus availability theory: Predictions and results for standard social contracts (from Cosmides, 1989, Experiments 1 and 2).

Familiarity, therefore, cannot account for the pattern of reasoning elicited by social contract problems. Social contract performance is not a by-product of familiarity.

Does Social Contract Content Simply Facilitate Logical Reasoning?

In the experiments just described, the adaptively correct answer if one is looking for cheaters happens to be the same as the logically correct answer—*P & not-Q*. Therefore, they cannot tell one whether performance on social contracts problems is governed by rules of inference that are specialized for reasoning about social exchange or by the rules of inference of the propositional calculus. Although we can think of no reason why this would be the case, perhaps social contract content simply facilitates logical reasoning. If so, then social contract performance could be a by-product of a logic faculty.

Two different sets of experiments show that this is not the case. The first involves "switched" social contracts (Cosmides, 1989, Experiments 3 and 4), and the second involves perspective change (Gigerenzer & Hug, in press).

Switched Social Contracts. The propositional calculus is content-independent: The combination of *P & not-Q* violates any conditional rule of the form *If P then Q*, no matter what "P" and "Q" stand for. The proposed social contract algorithms are not content-independent: Cheating is defined as accepting a benefit without paying the required cost. It does not matter what logical category these values happen to correspond to. For example, although the same social contract is expressed by both of the following statements, the proposition "you give me your watch" corresponds to the logical category *P* in the first rule and to *Q* in the second one.

Rule 1: "If you give me your watch, I'll give you $20" (standard form).
Rule 2: "If I give you $20, you give me your watch" (switched form).

No matter how the contract is expressed, I will have cheated you if I accept your watch but do not offer you the $20, that is, if I accept a benefit from you without paying the required cost. If you are looking for cheaters, you should therefore choose the "benefit accepted" card (I took your watch) and the "cost not paid" card (I did not give you the $20) no matter what logical category they correspond to. In the case of Rule 1, my taking your watch without paying you the $20 would correspond to the logical categories *P* and *not-Q,* which happens to be the logically correct answer. But in the case of Rule 2, my taking your watch without giving you the $20 corresponds to the logical categories *Q* and *not-P.* This is not the logically correct response. In this case, choosing the logically correct answer, *P & not-Q,* would constitute an adaptive error: If I gave you the $20 (*P*) but did not take your watch (*not-Q*), I have paid the cost but not accepted the benefit. This makes me an altruist or a fool, but not a cheater.

The general principle is illustrated in Figure 3.3, panel c, which shows the cost-benefit structure of a social contract. Rule 1 ("If you take the benefit, then you pay the cost") expresses the same social contract as Rule 2 ("If you pay the cost, then you take the benefit"). A person looking for cheaters should always pick the "benefit accepted" card and the "cost not paid" card. But for Rule 1, a "standard" social contract, these cards correspond to the logical categories *P* and *not-Q,* whereas for Rule 2, a "switched" social contract, they correspond to the logical categories *Q* and *not-P.* Because the correct social contract answer is different from the correct logical answer for switched social contracts, by testing such rules we can see whether social contracts activate inference procedures of the propositional calculus, such as modus ponens and modus tollens, or inference procedures that are specialized for detecting cheaters on social contracts.

The design of the following experiments was similar to that just described. Each subject solved four Wason selection tasks, presented in counterbalanced order: an unfamiliar social contract, an unfamiliar descriptive problem, a familiar descriptive problem, and an abstract problem. The only difference was that in this case the terms of the two unfamiliar rules were "switched" within the "If-then" structure of the rule. For example, instead of reading, "If you eat duiker meat, then you have found an ostrich eggshell," the rule would read, "If you have found an ostrich eggshell, then you eat duiker meat." This was true for both the unfamiliar social contract and the unfamiliar descriptive rule. For the social contract rules, this switch in the order of the terms had the effect of putting the cost term in the "If" clause, and the benefit term in the "then" clause, giving the rule the structure of the switched social contract shown in Rule 2 of Figure 3.3, panel c.

The predictions of social contract theory and the availability theories are shown in Figure 3.5, along with the results of two experiments. *Not-P & Q* is an extremely rare response on the Wason selection task, but social contract theory predicts that it will be very common on switched social contracts. That is exactly what happened. In fact, as many people chose *not-P & Q* on the unfamiliar switched social contracts as chose *P & not-Q* on the standard social contract problems described above.

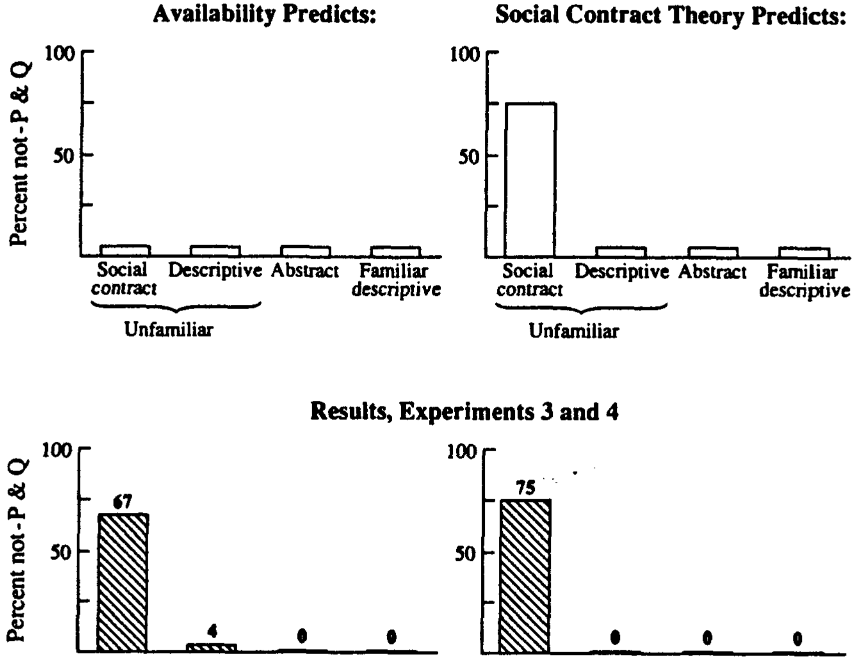


Figure 3.5 Social contract theory versus availability theory: Predictions and results for switched social contracts (from Cosmides, 1989, Experiments 3 and 4).

If social contract content merely facilitates logical reasoning, then subjects should have chosen *P & not-Q* on these switched social contract problems. The fact that they chose *not-P & Q*—a logically incorrect response—shows that social contract performance is not caused by the activation of a logic faculty. This is, however, the response one would expect if humans have rules of inference that are specialized for cheater detection.

Table 3.2 shows individual card choices for matching sets of experiments with standard versus switched social contracts, sorted by logical category and by social contract category. The results for non-social contract problems replicate beautifully when sorted by logical category. But not the results for the social contract problems. These results replicate when sorted by social contract category, not by logical category. This shows that the content-dependent social contract categories, not the logical categories, are psychologically real for subjects solving social contract problems. This confirms another predicted design feature of the social contract algorithms: They define cheating in terms of cost-benefit categories, not in terms of logical categories.

Manktelow and Over (1987) have pointed out that even when words such as "must" or "may" are left out of a social contract, one tends to interpret a standard social contract as meaning, "If you take the benefit, then you (must) pay the cost," whereas one tends to interpret a switched social contract as meaning, "If you pay the

Table 3.2 Selection Frequencies for Individual Cards, Sorted by Logical Category and by Social Contract Category[a]

	Unfamiliar Descriptive		Abstract Problem		Familiar Descriptive		Unfamiliar Social Contract	
							Standard	*Switched*
Logical Category	Exp. 1 & 2	Exp. 3 & 4	Exp. 1 & 2	Exp. 3 & 4	Exp. 1 & 2	Exp. 3 & 4	Exp. 1 & 2	Exp. 3 & 4
P	43	40	46	46	46	45	43	3
not-P	9	11	10	11	1	2	3	36
Q	20	23	15	23	8	6	0	44
not-Q	18	20	21	25	23	32	39	3
Social Contract Category:								
Benefit accepted							43	44
Benefit not accepted							3	3
Cost paid							0	3
Cost not paid							39	36

[a] Experiments 1 and 2 tested standard versions of the two unfamiliar rules, whereas Experiments 3 and 4 tested switched versions of these rules.

cost, then you (may) take the benefit." This is, in fact, a prediction of social contract theory: A cost is something one is obligated to pay when one has accepted a benefit, whereas a benefit is something that one is entitled to take (but need not) when one has paid the required cost. Thus, the interpretive component of the social contract algorithms should cause subjects to "read in" the appropriate "musts" and "mays," even when they are not actually present in the problem (three out of four of the standard social contracts had no "must," and none of the switched social contracts had a "may"). Could it be that subjects are in fact reasoning with the propositional calculus, but applying it to these reinterpretations of the social contract rules?

No. In the propositional calculus, "may" and "must" refer to possibility and necessity, not to entitlement and obligation. On the Wason selection task, the logically correct answer for the rule, "If you pay the cost, then it is possible for you to take the benefit," is to choose no cards at all. Because this rule admits only of possibility, not of necessity, no combination of values can falsify it. The fact that most subjects chose *not-P & Q*, rather than no cards at all, shows that they were not applying the propositional calculus to a rule reinterpreted in this way.[13] To choose *not-P & Q*, one would have to be following the implicational structure of social exchange specified in Cosmides and Tooby (1989).

Perspective Change. Gigerenzer and Hug (in press) have conducted an elegant series of experiments that test another design feature of the proposed social contract algorithms, while simultaneously showing that the results cannot be explained by the propositional calculus or by permission schema theory, which is discussed later. They gave two groups of subjects Wason selection tasks in which they were to look for violations of social contract rules such as, "If an employee gets a pension, then that employee must have worked for the firm for at least 10 years." The only difference between the two groups was that one group was told "You are the employer" whereas the other group was told "You are the employee."

In social contract theory, what counts as cheating depends on one's perspective. Providing a pension is a cost that the employer incurs to benefit the employee, whereas working 10 or more years is a cost that the employee incurs to benefit the employer. Whether the event "the employee gets a pension" is considered a cost or a benefit therefore depends on whether one is taking the perspectivc of the employer (= cost) or the employee (= benefit).The definition of cheating as taking a benefit without paying the cost is invariant across perspectives, but the theory predicts that which events count as benefit and cost will differ across actors. From the employer's perspective, cheating is when an employee gets a pension (the employee has taken the benefit) but has not worked for the firm for at least 10 years (the employee has not paid the cost). These cards correspond to the logical categories *P & not-Q*. From the employee's perspective, cheating is when an employee has worked for at least 10 years (the employer has taken the benefit), but has not been given the pension that he or she is therefore entitled to (the employer has not paid the cost). These cards correspond to the logical categories *not-P & Q*.

In other words, there are two different states of affairs that count as cheating in this situation, which correspond to the perspectives of the two parties to the exchange. What counts as cheating depends on what role one occupies; cheating is a well-defined concept, but its definition is preeminently content- and context-dependent.

In contrast, whether one is cued into the role of employer or employee is irrelevant to the content-independent propositional calculus. The correct answer on such a problem is *P & not-Q* (employee got the pension but did not work 10 years), regardless of whether the subject is assigned the role of the employer or the employee.

Gigerenzer and Hug conducted four experiments using different social contract rules to test the perspective change hypothesis. The predictions of both social contract theory and the propositional calculus are shown in Figure 3.6, along with the results. The results are as social contract theory predicts: Even though it is logically incorrect, subjects answer *not-P & Q* when these values correspond to the adaptively correct "look for cheaters" response. The hypothesis that social contract content simply facilitates logical reasoning cannot explain this result. The perspective change results and the switched social contract results show that social contract performance is not a byproduct of the activation of a logic faculty.

In light of these results, it is interesting to note that although schizophrenic individuals often perform more poorly than normals on problems requiring logical reasoning, deliberation, or seriation, Maljković found that their reasoning about social contracts is unimpaired (Maljković, 1987). She argues that this result makes sense if one assumes that the brain centers that govern reasoning about social contracts are different from those that govern logical reasoning.

To show adaptation, one must both eliminate by-product hypotheses *and* show evidence of special design for accomplishing an adaptive function. Algorithms specialized for reasoning about social exchange should have certain specific design features, and the switched social contract and perspective change experiments confirm three more predictions about those design features:

1. The definition of cheating embodied in the social contract algorithms should depend on one's perspective. The perspective change experiments confirm the existence of this design feature.
2. Computing a cost-benefit representation of a social contract from one party's

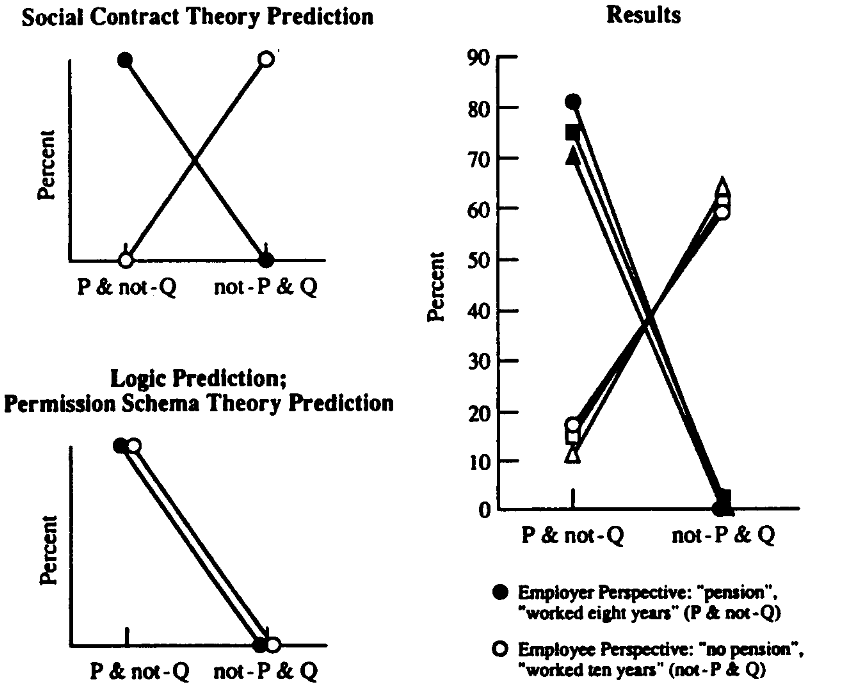


Figure 3.6 Perspective change experiments. Social contract theory versus logic facilitation hypothesis and permission schema theory: Predictions and results. Three separate experiments were conducted, testing the predictions of the theories against one another. The results of the three experiments are indicated by circles, squares, and triangles. Filled versus unfilled represents versions otherwise identical, except that the perspective (e.g., employer versus employee) is reversed (from Gigerenzer & Hug, in press).

perspective should be just as easy as computing it from the other party's perspective. There are two reasons for this prediction: First, to successfully negotiate with others, one must be able to compute the conditions under which others would feel that you had cheated them, as well as the conditions under which they had cheated you. Second, being able to understand what counts as cheating from both perspectives facilitates social learning; by watching other people's exchanges one can gather information about the values, and perhaps about the trustworthiness, of people one may interact with in the future. If people are just as good at translating a social contract into the values of one party as the other, then they should be just as good at detecting cheaters from one perspective as the other: There should be just as many *P & not-Q* responses to the "employer" version as there are *not-P & Q* responses to the "employee" version. This was, in fact, the case.

3. The implicational structure of social contract theory mandates that the statement "If an employee gets a pension, then that employee must have worked for the firm for at least 10 years" be taken to imply "If the employee has worked for 10 years, then the employer must give that employee a pension." This is because when the employee has fulfilled his obligation to benefit the employer, the employer is obligated to benefit the employee in return. If subjects draw this implication, then they should choose *not-P & Q* in both the switched social contract experiments and in the employee version of the perspective change experiments. The fact that they did confirms this design feature of the proposed social contract algorithms.

Is There a Cheater Detection Procedure, or Are People Simply Good at Reasoning about Social Contracts?

Social contract theory posits that the mind contains inference procedures specialized for detecting cheaters and that this explains the high percentage of correct social contract answers that these problems elicit. But maybe social contract problems simply "afford" clear thinking. Perhaps they are interesting, or motivationally compelling, in a way that other problems are not. Rather than having inference procedures specialized for detecting cheaters, perhaps we form a more complete mental model of a problem space for social contract problems, and this allows us to correctly answer *any* question that we might be asked about them, whether it is about cheating or not (for descriptions of mental model theories of reasoning, see Evans, 1984; Johnson-Laird, 1983; Manktelow & Over, 1987). Although it would be difficult to reconcile the perspective change data with this hypothesis, it is still worth considering.

No one has presented any independent criteria for judging what kinds of problems ought to be "interesting," "motivationally compelling," or "easy to understand," which makes this hypothesis nebulous. Nevertheless, it can be tested by studying performance on reasoning problems in which the rule is portrayed as a social contract, but in which the subject is *not* asked to look for cheaters.

Two sets of experiments did just that. One set asked subjects to look for altruists in a situation of social exchange; the other asked subjects to look for violations of a social contract rule in a context in which looking for violations did not correspond to looking for cheaters. If people are good at detecting cheaters merely because social contract problems are easy to understand, then performance on such problems should be just as good as performance on cheater detection problems.[14] But if people are good at detecting cheaters because they have inference procedures specialized for doing so, then such problems should elicit lower levels of performance than social contract problems that require cheater detection.

Are People Good at Looking for Altruists? The game-theoretic models for the evolution of cooperation that could be reasonably applied to the range of population structures that typified hominid hunter-gatherers require the existence of some mechanism for detecting cheaters or otherwise excluding them from the benefits of cooperation. This is because the capacity to engage in social exchange could not have evolved in the first place unless the individuals involved could avoid being continually exploited by cheaters. But most models do not require the existence of a mechanism for detecting "altruists"—individuals who follow the strategy of paying the required cost (thereby benefiting the other party), but not accepting from the other party the benefit to which this act entitles them.[15] Indeed, because individuals who were consistently altruistic would incur costs but receive no compensating benefits, under most plausible scenarios they would be selected out. Because they would not be a long-enduring feature of the adaptive landscape, there would be no selection pressure for "altruist detection" mechanisms. Thus, while we did expect the existence of inference procedures specialized for detecting cheaters, we did not expect the existence of inference procedures specialized for detecting altruists.

In contrast, if people are good at detecting cheaters merely because social contract problems afford clear thinking, then performance on altruist detection problems should be just as good as performance on cheater detection problems. The mental model of the social contract would be the same in either case; one would simply search the problem space for altruists rather than for cheaters.

To see whether this was so, we tested 75 Stanford students on some of the same social contract problems that were used by Cosmides (1985, 1989), but instead of asking them to look for cheaters, they were asked to look for altruists (the procedure was the same as that described in Cosmides, 1989, for Experiments 5 through 9). Each subject was given one social contract problem, and this problem required altruist detection. There were three conditions with 25 subjects each; a different social contract problem was tested in each condition. The first two conditions tested problems that portrayed a private exchange between "Big Kiku" (a headman in a fictitious culture) and four hungry men from another band; the third condition tested a problem about a social law. The first group's problem was identical to the cheater detection problem tested by Cosmides (1989) in Experiment 2, except for the instruction to look for altruists (see Figure 3.7 for a comparison between the cheater detection version and the altruist detection version). The problem in the second condition was essentially the same, but instead of portraying Big Kiku (the potential altruist) as a ruthless character, he was portrayed as having a generous personality. The third condition tested the social law, "If you eat duiker meat, then you have found an ostrich eggshell" (see Cosmides, 1989, Experiment 1). The instructions on this problem were suitably modified to ask subjects to look for altruists rather than cheaters.[16]

Because altruists are individuals who have paid the cost but have not accepted the benefit, subjects should choose the "benefit not accepted" card (*not-P*) and the "cost paid" card (*Q*) on these problems. These values would correspond to the "no tattoo" card and the "Big Kiku gave him cassava root" card for the first two problems, and to the "does not eat any duiker meat" card and the "has found an ostrich eggshell" card for the social law problem. Table 3.3 shows that the percentage of subjects who made this response was quite low for all three altruist detection problems.

Is it possible that Stanford students simply do not know the meaning of the word "altruistic"? We thought this highly unlikely, but just to be sure, we ran another 75 Stanford students ($n = 25$ per condition) on problems that were identical to the first three, except that the word "selflessly" was substituted for the word "altruistically." The word "selfless" effectively announces its own definition—less for the self. If there are inference procedures specialized for detecting altruists—or if social contract problems merely afford clear thinking—then surely subjects should be able to perform as well on the "selfless" problems as they do on cheater detection problems.

Table 3.3 Altruist Detection: Percent Correct *(not-P & Q)*

	Personal Exchange		Law
	Ruthless	Generous	
Altruistic	28	8	8
Selfless	40	36	12

Table 3.3 shows that this was not the case. Although performance was a bit higher on the problems that used "selfless" than on the problems that used "altruistic," performance was nowhere near the average of 74% that Cosmides (1989) found for comparable cheater detection problems.[17] In fact, performance on the selfless versions of the altruist detection problems was no better than performance on the familiar descriptive transportation problem (reported earlier). This indicates that people do not have inference procedures specialized for detecting altruists on social contracts, which is just what social contract theory predicted. More important, it casts doubt on the hypothesis that cheater detection problems elicit high levels of performance merely because social contracts afford clear thinking.

Are People Good at Looking for Violations of Social Contracts When These Do Not Indicate Cheating? Gigerenzer and Hug (in press) conducted a series of experiments designed to disentangle the concept of cheater detection from the concept of a social contract. The opportunity to illicitly benefit oneself is intrinsic to the notion of cheating. But one can construct situations in which the reason one is looking for violations of a social contract rule has nothing to do with looking for individuals who are illicitly benefiting themselves—i.e., one can construct situations in which looking for violations is *not* tantamount to looking for cheaters. Gigerenzer and Hug gave subjects Wason selection tasks in which all rules were framed as social contracts, but which varied in whether or not looking for violations constituted looking for cheaters.

Here is an example using the rule, "If one stays overnight in the cabin, then one must bring a load of firewood up from the valley." In the "cheating" version, it is explained that two Germans are hiking in the Swiss Alps and that the local Alpine Club has cabins at high altitudes that serve as overnight shelters for hikers. These cabins are heated by firewood, which must be brought up from the valley because trees do not grow at this altitude. So the Alpine Club has made the (social contract) rule, "If one stays overnight in the cabin, then one must bring along a bundle of firewood from the valley." There are rumors that the rule is not always followed. The subject is cued into the perspective of a guard whose job is to check for violations of the rule. In this version, looking for violations of the rule is the same as looking for cheaters.

In the "no cheating" version, the subject is cued into the perspective of a member of the German Alpine Association who is visiting a cabin in the Swiss Alps and wants to find out how the local Swiss Alpine Club runs the cabin. He sees people carrying loads of firewood into the cabin, and a friend suggests that the Swiss might have the same social contract rule as the Germans—"If one stays overnight in the cabin, then one must bring along a bundle of firewood from the valley." The story also mentions an alternative explanation: that members of the Swiss Alpine Club (who do not stay overnight in the cabin) bring wood, rather than the hikers.To settle the question, the subject is asked to assume that the proposed social contract rule is in effect, and then to look for violations of it. Note that the intent here is not to catch cheaters. In this situation, violations of the proposed social contract rule can occur simply because the Swiss Alpine Club never made such a rule in the first place.

In both versions, the rule in question is a social contract rule—in fact, exactly the same social contract rule. And in both versions, the subject is asked to look for violations of that rule. But in the cheating version, the subject is looking for violations

Cheater version:

You are an anthropologist studying the Kaluame, a Polynesian people who live in small, warring bands on Maku Island in the Pacific. You are interested in how Kaluame "big men" – chieftains – wield power.

"Big Kiku" is a Kaluame big man who is known for his ruthlessness. As a sign of loyalty, he makes his own "subjects" put a tattoo on their face. Members of other Kaluame bands never have facial tattoos. Big Kiku has made so many enemies in other Kaluame bands, that being caught in another village with a facial tattoo is, quite literally, the kiss of death.

Four men from different bands stumble into Big Kiku's village, starving and desperate. They have been kicked out of their respective villages for various misdeeds, and have come to Big Kiku because they need food badly. Big Kiku offers each of them the following deal:

"If you get a tattoo on your face, then I'll give you cassava root."

Cassava root is a very sustaining food which Big Kiku's people cultivate. The four men are very hungry, so they agree to Big Kiku's deal. Big Kiku says that the tattoos must be in place tonight, but that the cassava root will not be available until the following morning.

You learn that Big Kiku hates some of these men for betraying him to his enemies. You suspect he will cheat and betray some of them. Thus, this is a perfect opportunity for you to see first hand how Big Kiku wields his power.

The cards below have information about the fates of the four men. Each card represents one man. One side of a card tells whether or not the man went through with the facial tattoo that evening and the other side of the card tells whether or not Big Kiku gave that man cassava root the next day.

Did Big Kiku get away with cheating any of these four men? Indicate only those card(s) you definitely need to turn over to see if Big Kiku has broken his word to any of these four men.

A. got the tattoo

B. Big Kiku gave him nothing

C. no tattoo

D. Big Kiku gave him cassava root

Altruist version:

You are an anthropologist studying the Kaluame, a Polynesian people who live in small, warring bands on Maku Island in the Pacific. You are *interested in how Kaluame "big men" – chieftains – wield power.*

"Big Kiku" is a Kaluame big man who is known for his ruthlessness. As a sign of loyalty, he makes his own "subjects" put a tattoo on their face. Members of other Kaluame bands never have facial tattoos. Big Kiku has made so many enemies in other Kaluame bands, that being caught in another village with a facial tattoo is, quite literally, the kiss of death.

Four men from different bands stumble into Big Kiku's village, starving and desperate. They have been kicked out of their respective villages for various misdeeds, and have come to Big Kiku because they need food badly. Big Kiku offers each of them the following deal:

"If you get a tattoo on your face, then I'll give you cassava root."

Cassava root is a very sustaining food which Big Kiku's people cultivate. The four men are very hungry, so they agree to Big Kiku's deal. Big Kiku says that the tattoos must be in place tonight, but that the cassava root will not be available until the following morning.

You learn that Big Kiku hates some of these men for betraying him to his enemies. You suspect he will cheat and betray some of them. However, you have also heard that Big Kiku sometimes, quite unexpectedly, shows great generosity towards others – that he is sometimes quite altruistic. Thus, this is a perfect opportunity for you to see first hand how Big Kiku wields his power.

The cards below have information about the fates of the four men. Each card represents one man. One side of a card tells whether or not the man went through with the facial tattoo that evening, and the other side of the card tells whether or not Big Kiku gave that man cassava root the next day.

Did Big Kiku behave altruistically towards any of these four men? Indicate only those card(s) you definitely need to turn over to see if Big Kiku has behaved altruistically towards any of these four men.

A. got the tattoo

B. Big Kiku gave him nothing

C. no tattoo

D. Big Kiku gave him cassava root

Figure 3.7 Both problems describe a social contract rule, but the problem on the left asks the subject to look for cheaters (individuals who took the benefit without paying the cost), whereas the problem on the right asks the subject to look for altruists (individuals who paid the cost but did not take the benefit to which this entitles them).

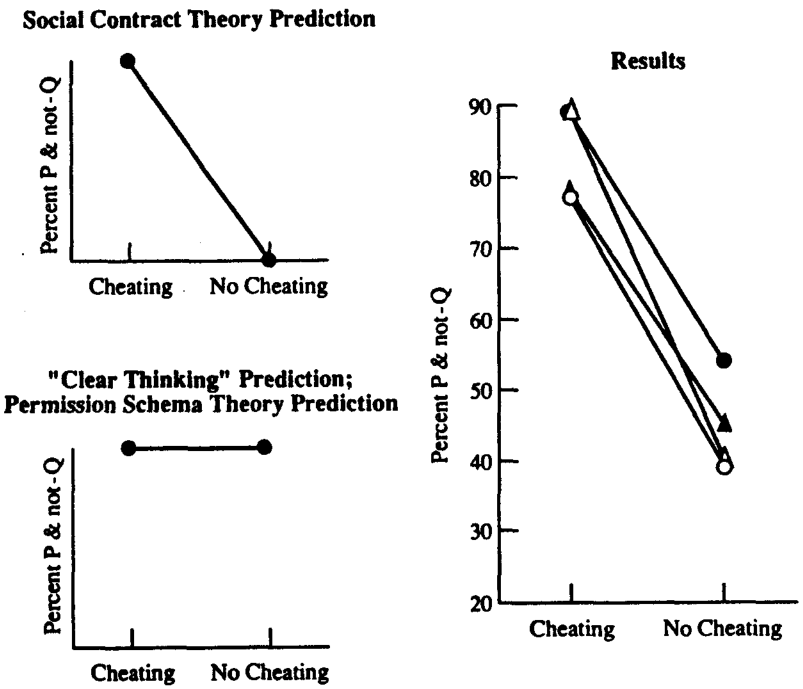


Figure 3.8 Is cheating necessary? Predictions and results. In these experiments, both conditions tested social contract rules, but in the "no cheating" condition looking for violations did not correspond to looking for cheaters. Social contract theory predicts a different pattern of results from both permission schema theory and the hypothesis that social contract content merely *affords* "clear thinking." Four separate experiments were conducted, testing the predictions of the theories against one another. The results of the four experiments are indicated by filled and unfilled triangles, and filled and unfilled circles (from Gigernzer & Hug, in press).

because he or she is looking for individuals who are illicitly benefiting themselves, whereas in the no cheating version the subject is looking for violations because he or she is interested in whether the proposed social contract rule is in effect. If social contract problems merely afford clear thinking, then a violation is a violation is a violation: It shouldn't matter whether the violation constitutes cheating or not.[18] In contrast, if there are inference procedures specialized for cheater detection, then performance should be much better in the cheating version, where looking for violations is looking for illicitly taken benefits. Figure 3.8 shows the predictions and the results of four such experiments. On average, 83% of subjects correctly solved the cheating version, compared with only 44% on the no cheating version. Cosmides and Tooby (in prep. b) conducted a similarly designed experiment, with similar results. In both problems, it was stipulated that a social contract rule was in effect and that the people whose job it was to enforce the rule may have violated it. But in one version violations were portrayed as due to cheating, whereas in the other version violations were portrayed as due to innocent mistakes. We found 68% of subjects correctly solved the cheating version, compared with only 27% of subjects on the "mistake" (no cheating) version. Thus, using a different "no cheating" context in which it was stipulated that the social contract rule was in effect, we were able to replicate the difference that Gigerenzer and Hug found between their cheating and no cheating versions almost exactly (39 percentage points in Gigerenzer and Hug; 41 percentage points in Cosmides and Tooby).

These data indicate that social contract problems do not merely afford clear thinking. In addition, these results provide further evidence against the availability theories and the hypothesis that social contract content merely facilitates logical reasoning.

The results of these experiments are most parsimoniously explained by the assumption that people have inference procedures specialized for detecting cheaters: individuals who have illicitly taken benefits. Because these procedures operate on the cost-benefit representation of a problem, they can detect a violation only if that violation is represented as an illicitly taken benefit. They would not be able to detect other kinds of violations, nor would they detect altruists. These results provide further confirmation of a proposed design feature of the social contract algorithms: This bundle of cognitive processes appears to include inference procedures specialized for cheater detection.

Are Cost-Benefit Representations Necessary, or Are Subjects Good at Detecting Violations of Any Rule Involving Permission?

The social contract algorithms should contain decision rules that govern when one should engage in an exchange. A proposed contract should not be accepted unless the costs one incurs by providing a benefit to the other party are outweighed by the benefits one receives by engaging in the exchange. Consequently, the first stage of processing must be the assignment of cost-benefit weightings to the "benefit term"—the event, item, or state of affairs the contract entitles one to receive—and to the "cost term"—the event, item, or state of affairs that the other party requires of one. We have been calling these terms the "benefit" term and the "cost" term for ease of explication; we do not mean to prejudge the actual values that the parties to the exchange assign to these terms. For example, satisfying the requirement may actually benefit the person who satisfies it—the offerer's belief that the benefit she is requiring the other party to provide represents a cost to that party may be erroneous. But the offerer would not propose the contract in the first place unless she believed that the state of affairs she wants to bring about would not occur in the absence of the contract—it would be silly to offer you $20 for your watch if I believed you were going to give it to me anyway. Similarly, the offerer may be mistaken in her belief that she is offering a benefit to the other party, in which case her offer will not be accepted—you might value your watch more than you would value having an extra $20. In social contract theory, costs and benefits are defined with respect to the value systems of the participants, and relative to a zero-level utility baseline that represents how each party would behave in the absence of the contract. (The cost-benefit conditions that should adhere for an offerer to offer and the acceptor to accept a social contract are spelled out in Cosmides, 1985, and Cosmides & Tooby, 1989.)

Once the contract has been translated into cost-benefit terms, the decision rules for acceptance or rejection can operate on that representation. The cheater detection procedures should also act on that representation, causing one to investigate individuals who have not fulfilled the requirement specified in the cost term and individuals who have taken the benefit offered. By operating on this relatively abstract level of representation, the cheater detection procedures can detect cheaters no matter what the actual items of exchange are. Because humans are a technological species capable of exchanging a wide array of items, we hypothesized that human cheater detection procedures would operate on this abstract cost-benefit level of representation, rather than be tied to any particular item of exchange, as discussed earlier in this chapter. In other species, the constraint that cheaters must be detected might be implemented by algorithms that operate on other levels of representation (as discussed for the vampire bat).

Thus, a predicted design feature of human social contract algorithms is that they cause one to assign cost-benefit representations to a social contract, one for each party to the interaction. The presence of this design feature was confirmed by the perspective change experiments. Another predicted design feature is that cheater detection procedures operate on cost-benefit representations. Indeed, it is difficult to see how they could do otherwise, given that cheating is defined as accepting a benefit one is not entitled to. Technically, a violation that does not illicitly benefit one is not cheating.

If cheater detection algorithms operate on cost-benefit representations, then they should not be able to operate on rules that are similar to social contracts but that have not been assigned a cost-benefit representation. To test this prediction, we investigated Wason selection task performance on permission rules that did not afford the assignment of a cost-benefit representation.

Permission rules are prescriptive rules that specify the conditions under which one is allowed to take some action. Cheng and Holyoak (1985) define them as "regulations . . . typically imposed by an authority to achieve some social purpose" (p. 398). Their "permission schema theory" proposes that people have a schema for reasoning about permissions that consists of the following four production rules:

Rule 1: If the action is to be taken, then the precondition must be satisfied.
Rule 2: If the action is not to be taken, then the precondition need not be satisfied.
Rule 3: If the precondition is satisfied, then the action may be taken.
Rule 4: If the precondition is not satisfied, then the action must not be taken.

On a Wason selection task in the linguistic format of rule 1, the first production rule would cause one to choose the "action has been taken" card (*P*), and the fourth production rule would cause one to choose the "precondition has not been satisfied" card (*not-Q*). Rules 2 and 3 would not cause any card to be chosen. If a Wason selection task presented a rule in the linguistic format of rule 3 (a cognate to our switched social contracts), then the same two cards would be chosen for the same reason, but they would now correspond to the logical categories *Q* and *not-P*.

Cheng and Holyoak (1985) and Cheng, Holyoak, Nisbett, and Oliver (1986) also propose the existence of "obligation schemas." These have the same implicational structure as permission schemas (and hence lead to the same predictions), but their representational format is "If condition C occurs, then action A must be taken." Because in any concrete case it is difficult to tell a permission from an obligation, we will refer to both kinds of rules under the rubric of permission schema theory.[19]

All social contracts are permission rules, but not all permission rules are social contracts. Social contract rules that have the form "If one takes the benefit, then one must pay the cost" are subsets of the set of all permission rules, because taking a benefit is just one kind of action that a person can take. Taking a benefit always entails taking an action, but there are many situations in which taking an action does not entail taking a benefit.

Permission schema theory has already been falsified by the experiments of Gigerenzer and Hug, described earlier. In permission schema theory, a permission rule has been violated whenever the action has been taken but the precondition has not been satisfied. It should not matter *why* subjects are interested in violations. Whether one is interested in violations because one is interested in finding cheaters or because one is interested in seeing whether the rule is in effect is irrelevant to permission schema theory. As long as the subject recognizes that the rule is a permission rule, the permission schema should cause the "action taken" and the "precondition not met" cards to be chosen. Consequently, permission schema theory would have to predict equally high levels of performance for the cheating and the no cheating versions of the social contract rules tested by Gigerenzer and Hug, as Figure 3.8 shows. Yet, even though both problems involved looking for violations of the same (social contract) permission rule, the cheating version elicited much higher performance than the no cheating version. Gigerenzer and Hug's perspective change experiments also falsify permission schema theory, and for a similar reason. Only one combination of values—"action taken" and "precondition not satisfied"—violates a permission rule in permission schema theory. What counts as a violation does not change depending on one's perspective; indeed, permission schema theory has no theoretical vocabulary for discussing differences of perspective. Yet the subjects' definition of violation depended on what role they were cued into, the employer's (*P & not-Q*) or the employee's (*not-P & Q*). Permission schema theory can account for the *P & not-Q* response, but not for the *not-P & Q* response.[20]

Permission schema theory and social contract theory differ in yet another way: The permission schema operates on the more abstract and inclusive action-precondition level of representation, whereas social contract algorithms construct and operate on the somewhat less general cost-benefit level of representation. Consequently, permission schema theory predicts that all permission rules will elicit high levels of performance, whether they have the cost-benefit structure of a social contract or not. In contrast, social contract theory does not predict high levels of performance for permission rules that do not afford the assignment of a cost-benefit representation. This is because cheating is *defined* as an illicitly taken benefit; where there are no benefits, there can be no cheating. By comparing performance on permission rules that do and do not afford the assignment of the appropriate cost-benefit representation, we can both test the prediction that cheater detection algorithms require a cost-benefit representation to operate and provide yet another test between social contract theory and permission schema theory.

In the first series of experiments (reported in Cosmides, 1989), we used the same research strategy as before: Wason selection tasks using rules whose terms and relation would be unfamiliar to our subjects, surrounded by a story context that either did or did not afford the assignment of the cost-benefit structure of a social contract to the rule. But in these experiments, the non-social-contract rule was a permission rule, not a descriptive rule. Let us illustrate the difference with an example: the school rule.

The school rule tested in both conditions was "If a student is to be assigned to Grover High School, then that student must live in Grover City." In the social contract version, the story explained that Grover High School is a much better high school than Hanover High, with a good record for getting students into college. Citizens of Grover City pay higher taxes for education than citizens of the town of Hanover, which is why Grover High is the better school. The story surrounding the rule thus portrays going to Grover High as a rationed benefit that must be paid for through higher taxes. Volunteers, some of whom are mothers with high-school-age children, are processing the school assignment documents, and it is rumored that some might have cheated on the rule in assigning their own children to a school. The subject is cued into the role of someone who is supervising these volunteers and must therefore look for cheaters.

In the non-social-contract version, the same permission rule is used, but the surrounding story did not afford a cost-benefit interpretation of the terms of the rule. Grover High is not portrayed as any better than Hanover High, nor does the story mention any greater cost that is incurred by living in Grover City rather than Hanover. It is, however, explained that it is important that this rule for assigning students from various towns to the appropriate school district be followed, because the population statistics they provide allow the board of education to decide how many teachers need to be assigned to each school. If the rule is not followed, some schools could end up with too many teachers and others with too few. Thus the story context gives the rule a "social purpose." The subject is cued into the role of a person who is replacing the absent-minded secretary who was supposed to follow this rule in sorting the students' documents. Because the former secretary frequently made mistakes, the subject must check the documents to see if the rule was ever violated.

Although this rule is a permission rule—stating the conditions under which one is allowed to assign a student to Grover High School—nothing in the story affords the assignment of the cost-benefit structure of a social contract to this rule. There is no benefit to assigning someone to Grover rather than to Hanover High, and no cost associated with living in one city rather than the other. Moreover, there is no apparent way that the absent-minded secretary could have illicitly benefited from breaking the rule. Thus her mistakes would not constitute cheating. By hypothesis, cheater detection algorithms should not be able to operate on this problem, because they would have no cost-benefit representations to attach themselves to. Social contract theory, therefore, predicts a lower percent of correct responses for this version than for the social contract version. In contrast, permission schema theory predicts high levels of the correct responses for both rules, because both are permission rules: Both have the action-precondition representational format that permission schema theory requires. (For both theories, the "correct" response is *P & not-Q* for standard rules and *not-P & Q* for switched rules.)

The predictions and results of four experiments—two with standard rules, two with switched rules—are displayed in Figure 3.9. Across the four experiments, 75% of subjects chose the correct answer for the social contract permission rules, compared with only 21% for the non-social-contract permission rules.

Using unfamiliar rules with a long story context has the advantage of giving the experimenter some control over the subject's mental model of the situation. The disadvantage of this method, however, is that it is difficult to create matching stories in which only one element varies. In the matched school rules, for example, two elements that distinguish permission schema theory and social contract theory varied: (a) whether the rule was given a cost-benefit structure, and (b) whether the potential violator was portrayed as a cheater or as a person who might have broken the rule by mistake. So we tackled the question of whether the rule must have a cost-benefit structure in another way as well: We tested minimalist problems that varied only in whether the subjects' past experience would cause them to interpret the antecedent of the rule as a benefit (Cosmides & Tooby, in prep., b). These problems had virtually no context: They simply explained that among the Kalama (another fictitious culture) the elders make the laws, and one of the laws they made is, "If one is going out at night, then one must tie a small piece of red volcanic rock around one's ankle." The subject was then asked to see if any of four individuals had violated this law. The rule was based on one developed by Cheng and Holyoak (1989), but the context gave the rule no rationale or "social purpose."

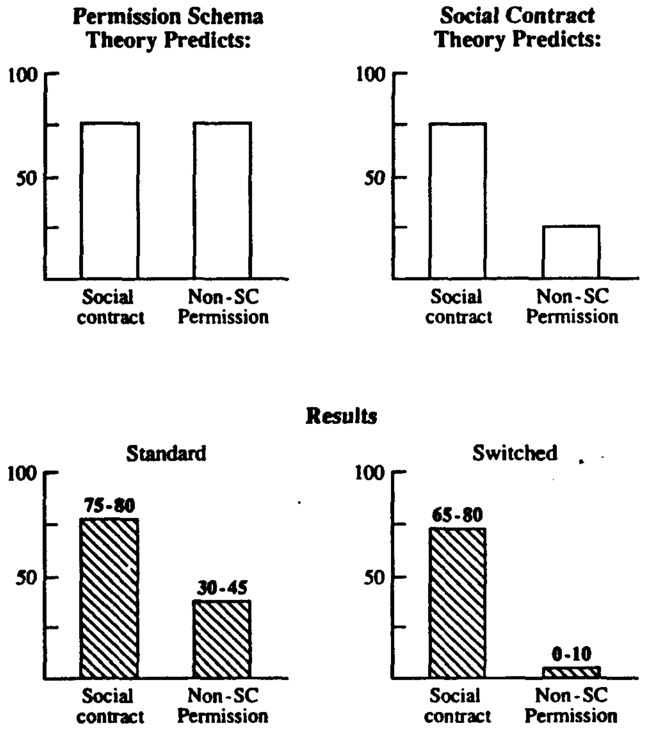


Figure 3.9 Are cost-benefit representations necessary? Social contract theory versus permission schema theory: Predictions and results (from Cosmides, 1989, Experiments 5, 6, 8, and 9).

We tested three versions of the rule that differed in only one respect: how much of a benefit the antecedent would seem to be. Among undergraduates, going out at night is a benefit: It represents fun, dating, adventure, and so on. Staying home at night is not as much fun, and taking out the garbage is even less so. Consequently, we compared performance on the "going out at night" rule to performance on two other rules: "If one is staying home at night, then one must tie a small piece of red volcanic rock around one's ankle" and "If one is taking out the garbage, then one must tie a small piece of red volcanic rock around one's ankle." If permission schema theory were correct, then all three of these permission rules would elicit equally high levels of *P & not-Q* responses.[21] But if social contract theory is correct, and the rule must have the cost-benefit structure of a social contract to elicit the effect, then performance should decline as the value of the antecedent declines. The more difficult it is to interpret the antecedent as a benefit, the more difficult it should be to see how one could illicitly benefit by breaking the rule.

The predictions and results are depicted in Figure 3.10. Performance decreases as the size of the benefit in the antecedent decreases, just as social contract theory predicts. Figure 3.10 also depicts the results of another, similar experiment with the so-called "Sears problem." As we removed the cost-benefit structure of the Sears problem, performance decreased. This experiment is also described in Cosmides and Tooby (in prep., b).

Manktelow and Over (1990; scenarios B and C) tested two obligation rules that lacked the cost-benefit structure of a social contract, with similar results. Cheng and Holyoak's theory predicts that both of these rules will elicit a high percentage of *P & not-Q* responses, yet they elicited this response from only 12% and 25% of subjects, respectively.

These experiments eliminate yet another by-product hypothesis: They show that reasoning on social contract problems is not a by-product of inference procedures for reasoning about a more general class of problems, permission problems. Permission rules elicit the effect only if the cost-benefit representation of a social contract can be assigned to them and if violating the rule would illicitly benefit the violator. These results confirm another predicted design feature: They show that cheater detection algorithms do not operate properly unless the appropriate cost-benefit representation can be assigned to the rule.

Will Any Rule That Involves the Possibility of Positive or Negative Payoffs Elicit Good Performance on the Wason Selection Task? Manktelow and Over (1990) were interested in this question when they designed the experiments just described. Their obligation rules lacked the cost-benefit structure of a social contract and, therefore, the property that one could illicitly benefit by cheating. But in both scenarios a person would receive high payoffs for following the rule. Violating the rule would cause a person to incur a cost; a small one in the case of one of the scenarios, a large one in the case of the other. They found that the possibility of payoffs, either positive or negative, was not sufficient to elicit high levels of *P & not-Q* responses. We found similar results in experiments testing rules concerning the possibility that a food was poisonous (Cosmides & Tooby, in prep., c), rules such as "If a person eats the red berries, then that person will vomit."

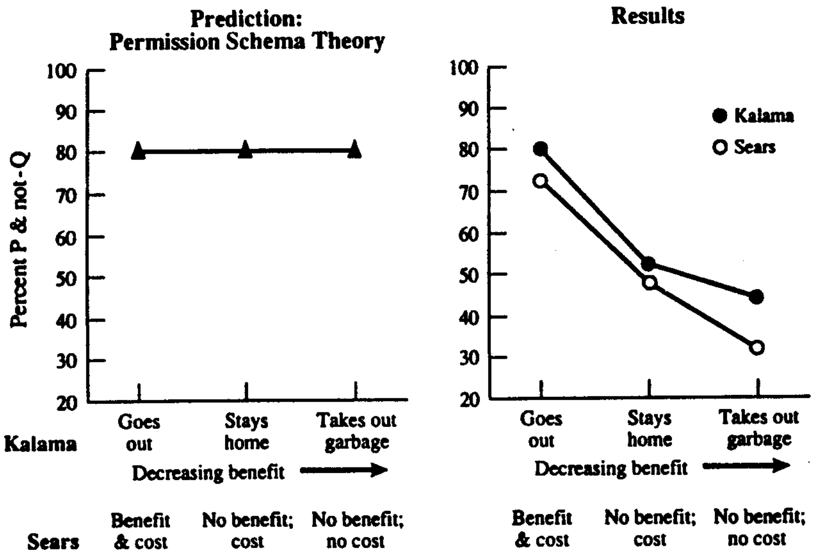


Figure 3.10 Are cost-benefit representations necessary? Social contract theory versus permission schema theory: Predictions and results (from Cosmides & Tooby, in prep., b).

These experiments eliminate another by-product hypothesis: They show that the possibility of a payoff is not sufficient to elicit good performance on the Wason selection task and, therefore, not sufficient to explain the social contract effect. To be good at detecting violations of a social contract rule, the violation has to represent an illicitly taken benefit.

The possibility of payoffs is not sufficient to explain the social contract effect. But heuristically, it is an important dimension to consider when one is trying to discover specialized inference procedures. An evolutionary approach would lead one to investigate domains in which our foraging ancestors would have experienced positive (fitness) payoffs for "correct" reasoning and negative (fitness) payoffs for "incorrect" reasoning. We place "correct" and "incorrect" in scare quotes because these notions are defined with respect to the adaptive problem to be solved; "correct" does not necessarily mean the logically correct *P & not-Q*, just as it did not in many of the social contract experiments reported here. In addition to social contracts, there is now experimental evidence suggesting the existence of specialized inference procedures for two other domains, both of which involve large fitness payoffs: precautions (prudential obligations) and threats. Manktelow and Over (1990; scenario A) were the first to uncover the possibility of precaution schemas for rules of the general form "If one enters a hazardous situation, then one should take the appropriate precaution," and Cheng and Holyoak (1989) and we (Cosmides & Tooby, in prep., b) have provided evidence in support of the existence of precaution schemas. But the experiments to be reported in Cosmides and Tooby (in prep., b) also show that precaution rules are processed differently from social contract rules, in accordance with a different adaptive logic appropriate to that problem. This is also true of threats. Our tests of threat rules show that people are very good at detecting bluffs and double crosses in situations of threat, but again, the reasoning involved does not map on to cheater detection in social contract situations (Cosmides & Tooby, in prep., a).

Summary of Experimental Findings

Virtually all the experiments reviewed above asked subjects to detect violations of a conditional rule. Sometimes these violations corresponded to detecting cheaters on social contracts, other times they did not. The results showed that we do not have a general-purpose ability to detect violations of conditional rules. But human reasoning is well designed for detecting violations of conditional rules when these can be interpreted as cheating on a social contract. Based on our computational theory of social exchange, we predicted that reasoning about social contracts would exhibit a number of specific design features. The results of the experiments reviewed above confirm the existence of many of these design features and do not falsify any of them. They show the following:

1. The algorithms governing reasoning about social contracts include inference procedures specialized for cheater detection.
2. Their cheater detection procedures cannot detect violations that do not correspond to cheating (such as mistakes).
3. The algorithms governing reasoning about social contracts operate even in unfamiliar situations.
4. The definition of cheating that they embody depends on one's perspective.
5. 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.
6. They cannot operate so as to detect cheaters unless the rule has been assigned the cost-benefit representation of a social contract.
7. They embody implicational procedures specified by the computation theory (e.g., "If you take the benefit then you are obligated to pay the cost" implies "If you paid the cost, then you are entitled to take the benefit").
8. They do not include altruist detection procedures.

Furthermore, the highly patterned reasoning performance elicited by social contracts cannot be explained as a by-product of any of the more general-purpose reasoning procedures that have been proposed so far. The following by-product hypotheses were eliminated:

1. That familiarity can explain the social contract effect.
2. That social contract content merely activates the rules of inference of the propositional calculus.
3. That social contract content merely promotes, for whatever reason, clear thinking.
4. That permission schema theory can explain the social contract effect.
5. That any problem involving payoffs will elicit the detection of violations.
6. That a content-independent formal logic, such as the propositional calculus, quantificational logic, or deontic logic can explain the social contract effect.

These findings strongly support the hypothesis that the human mind includes cognitive procedures that are adaptations for reasoning about social exchange.

IMPLICATIONS FOR CULTURE

Cultural Forms Are Structured by Our Universal, Evolved Psychology

Wherever human beings live, their cultural forms and social life are infused with social exchange relations (e.g., Malinowski, 1922; Mauss, 1925/1967). Such relations appear in an enormous range of different guises, both simple and highly elaborated, implicit and explicit, deferred and simultaneous, practical and symbolic. The magnitude, variety, and complexity of our social exchange relations are among the most distinctive features of human social life, and differentiate us strongly from all other animal species (Tooby & DeVore, 1987).

Antiquity, universality, and cross-cultural elaboration are exactly what one expects of behaviors that are the expression of universal, evolved information-processing mechanisms. From the child who gets dessert if her plate is cleaned, to the devout Christian who views the Old and New Testaments as covenants arrived at between humans and the supernatural, to the ubiquitous exchange of women between descent groups among tribal peoples, to trading partners in the Kula Ring of the Pacific—all of these phenomena require, from the participants, the recognition and comprehension of a complex set of implicit assumptions that apply to social contract situations. Our social exchange psychology supplies a set of inference procedures that fill in all these necessary steps, mapping the elements in each exchange situation to their representational equivalents within the social contract algorithms, specifying who in the situation counts as an agent in the exchange, which items are costs and benefits and to whom, who is entitled to what, under what conditions the contract is fulfilled or broken, and so on (Cosmides & Tooby, 1989).

Without social exchange and the underlying constellation of cognitive adaptations that support it, social and mental life in every culture would be so different as to be unrecognizable as *human* life. If one removed from our evolutionary history and hence from our minds the possibility of cooperation and reciprocity—of mutually contingent benefit-benefit interactions arrived at through mutual consent—then coercion and force would loom even larger as instruments of social influence, and positive relationships would be limited primarily to self-sacrificial interactions among kin. Such conditions do, in fact, typify the social life of most other animal species. What psychological mechanisms allow humans to behave so differently?

According to the Standard Social Science Model, "culture" builds all concepts as sophisticated and as content-specific as social exchange from scratch, using only content-free general-purpose mental processes (see Tooby & Cosmides, this volume). Yet the experiments reviewed herein have shown that no general-purpose process so far proposed can produce the sophisticated pattern of reasoning needed to engage in social exchange. Moreover, these experiments cast serious doubt on the claim that content-free general-purpose mechanisms were responsible for *building* the content-specific social contract algorithms. Although we have reasoning procedures elaborately tailored for solving this ancient adaptive problem, these experiments have demonstrated that we do not have reasoning procedures that are similarly specialized for solving many familiar problems that are routinely encountered in the modern social world—such as detecting whether someone has made a mistake, detecting whether someone has broken a prescriptive rule that is not a social contract, and detecting whether a situation exists that violates a descriptive rule. Yet, if we did have a content-free psychology that could build the necessary reasoning mechanisms "as needed," then one would expect to find elaborate reasoning procedures specialized for solving such problems. Procedures specialized for solving ancient adaptive problems, such as social exchange, would be no more likely to develop than procedures specialized for solving the many evolutionarily novel problems posed by life in a modern, postindustrial culture. Evaluating a scientific hypothesis about the effects of dietary cholesterol, detecting whether someone has misfiled a document, and following the string of "if-then" directions on one's tax returns would be as effortless—and require as little explicit instruction—as detecting cheaters in a situation of social exchange among members of an unfamiliar tribal culture who value cassava root and facial tattoos.

The claim that our only evolved psychological mechanisms are general-purpose and content-free, and that "culture" must therefore supply all the specific content of our minds, is exactly the issue on which evolutionary psychological approaches diverge most sharply from more traditional ones. In our view, instead of culture manufacturing the psychology of social exchange de novo, content-specific, evolved psychologies constitute the building blocks out of which cultures themselves are manufactured (Tooby & Cosmides, this volume). These psychologies evolved to process information about ancient and important adaptive problems, such as social exchange, sexual jealousy, kin recognition, language acquisition, emotion recognition, and parenting. On this view, the social environment need not provide all of the (hypothetical) properties needed to construct a social exchange psychology from a set of content-free mental procedures (assuming this is possible at all), because the evolved architecture of the human mind already contains content-specific mechanisms that will cause a social exchange psychology to reliably develop in every normal human being. Because every human being will develop the same basic set of social contract algorithms, cultural forms that require their presence can emerge: traditions, rituals, institutions, linguistic conventions, symbols, and so forth can develop that rely on the stable features of this psychology, and that simply supply the specifics that activate and deploy it in each new situation.

Thus, there is an immediate difference between evolutionary psychological approaches and approaches that maintain that everything involving social interaction is "socially constructed." An evolutionary psychological perspective suggests that the universal evolved architecture of the human mind contains some content-specific algorithms that are shared across individuals and across cultures, and that, therefore, many things related to social exchange should be the same from place to place—as indeed they are. In contrast, a Standard Social Science Model approach would, if thoroughgoing, maintain that social exchange is culture-specific and historically contingent, existing in some places and not in others. Moreover, the Standard Model would have to predict that wherever social exchange is found to exist, it would have to be taught or communicated from the ground up. Because nothing about social exchange is initially present in the psychology of the learner, every structural feature of social exchange must be specified by the social environment as against the infinity of logically alternative branchings that could exist. It is telling that it is just this explicitness that is usually lacking in social life. Few individuals are able to articulate the assumptions that structure their own cultural forms—to do so usually requires considerable effort and is accompanied by the awareness that one is doing something unusual and artificial (Sperber, 1975). We suggest that social exchange is learned without explicit enumeration of the underlying assumptions. The surface forms are provided by specific cultures, but they are interpreted by content-specific information-processing devices containing implicit assumptions and procedures that evolved for exactly this purpose.

The issue of explicitness is particularly instructive. As Sperber and Wilson (1986) point out, if two individuals have no shared assumptions about the world, communication between them is impossible. The explicit part of the communicative process concerns those features of a situation that are known to the sender but not yet known to the receiver, but such transmissions are only interpretable because of what is already mutually understood. This point is particularly relevant to understanding human culture, because the learning of a culture itself in effect consists of successful communication between members of the culture and the individual who is learning about it. Consequently, for someone ignorant of a culture to learn it in the first place, it is necessary that she already share many assumptions with those from whom she is learning the culture. If human minds truly were initially tabula rasas, with no prior contentful structure, then no anthropologist or immigrant to a culture could ever learn about it (Quine, 1969). More to the point, all children are immigrants,born into the world with no culturally specific knowledge at all. If the evolved architecture of the child's mind contained no meaningful content at all—that is, if it were truly content-free—then children could never learn the culture they are born into: They would have no reliable means of interpreting anything they saw and hence of placing the same construction on it that members of their culture do.

We suggest that domain-specific reasoning procedures such as social contract algorithms (Cosmides, 1989; Cosmides & Tooby, 1989) supply what is missing from traditional accounts of the acquisition of culture: that is, the necessary preexisting conceptual organization, which provides the shared assumptions humans need to interpret unfamiliar behavior and utterances and, in so doing, to acquire their culture. Such domain-specific algorithms embody both intrinsic definitions for the representations that they operate on (e.g., benefit, requirement, cheater) and cues for recognizing which elements in ontogenetically unfamiliar but evolutionarily recurrent situations correspond to those representations. This is why humans can barter or even form friendships where no common language is spoken. It is for this reason that we tested our subjects' ability to apply social contract algorithms to unfamiliar situations: Indeed, they did so as easily as they apply them to familiar contents. For all of the above reasons, we argue that such content-sensitive cognitive adaptations are a critical element in the process of cultural transmission. Culture would not be possible without them.

In any case, the way in which such a universal psychology would lead to cultural universals is straightforward, and the enumeration of cases of social exchange found cross-culturally would add little to furthering the discussion. For that reason, we would like to focus on two more interesting issues: (a) Assuming there is a universal, evolved psychology of social exchange, how might one explain and organize our understanding of cultural differences? That is, how can an underlying universal psychology express itself differently in different cultures? And (b) How can the ethnographic record inform us about additional aspects of our social exchange psychology, beyond those few algorithms and representations that have already been explored experimentally?

Intergroup Differences: Evoked Culture and Transmitted Culture

Human thought and behavior differs in an organized fashion from place to place, and these differences are typically termed "cultural" variation. In fact, it is considered a tautology by many to attribute this variation to the operation of culture—that is, to social transmission. Nevertheless, there are two complementary explanations for the existence of these within-location similarities and between-location differences in thought and behavior (Tooby & Cosmides, 1989, this volume). The traditional explanation is that these systematic differences are caused by what is usually called *transmitted culture,* that is, the process whereby the thought and behavior of some individuals (usually from the preceding generation) is passed on to other individuals, thereby causing the present pattern. Indeed, we have already touched on how domain-specific cognitive adaptations make cultural transmission possible.

There is, however, a second way in which such local similarities can be brought about. Our universal, evolved information-processing mechanisms should be sensitively context-dependent: Different informational inputs should evoke different representational and behavioral outputs. Humans share the same evolved information-processing mechanisms, but they do not all live under the same circumstances. People living in the same location are likely to experience somewhat similar circumstances, which should evoke the same kind of response from each individual; people living in different locations are likely to experience somewhat different circumstances, which should evoke different responses from each individual. To take an obvious example, people in tropical forest environments all around the world tend to be at most only lightly clothed, something that has less to do with parental example than with heat load.

The more complex the set of species-typical psychological mechanisms, the more sensitively dependent behavior may be on local circumstances. This means that cultural forms may exist that are not primarily generated by cultural transmission at all. We call similarities triggered by local circumstances *evoked culture* (Tooby & Cosmides, this volume). Of course, these two explanations are by no means mutually exclusive and usually operate together to shape the distribution of human similarities and differences. The evolutionary psychology of social exchange shapes both evoked culture and transmitted culture, and the exploration of a few of these issues serves to illustrate how an evolutionary psychological approach to culture can diverge from traditional anthropological approaches.

Open Questions in the Psychology of Social Exchange

Before proceeding to discuss how an evolved, universal psychology can generate cultural variation, it is necessary to emphasize the limited nature of what we have done so far. We have tested only a small number of hypotheses about one mode of social exchange out of many. We have not discussed, much less experimentally explored, the rest of the intricate and complex psychology of social exchange. Anyone examining his or her own human experience will immediately identify large areas of the psychology of social exchange, such as the psychology of friendship, that are not captured well by the models introduced so far. More important, the other components of our evolved faculty of social cognition—for example, the psychological mechanisms that govern sexual relations, coalitional partnerships, status, revenge, threat, and parenting—will have to be mapped out and integrated with the psychological mechanisms governing social exchange before social exchange can be fully unraveled. Each component of the faculty of social cognition can be only incompletely understood when taken in isolation.

The computational theory that we have developed specifies certain contractual relationships and implicit inferences that people can be expected to make in social contract situations (Cosmides & Tooby, 1989). It is, however, very far from a complete computational theory of social exchange and should not be mistaken for one: Instead, it is only an initial exploration of some of its major features. Many crucial questions were left unaddressed—questions that must be answered if we are to understand social exchange in all its various forms across cultures, or even within a single culture. For example, how many instances of cheating should the decision rule tolerate before it activates mechanisms causing one to sever one's relationship with an individual? Under what conditions should one cooperate with a person on a short-term basis, as opposed to a long-term basis? Should one's willingness to tolerate cheating differ in short-term versus long-term relationships? Will variance in the acquisition of the resources being exchanged affect the ways in which they are shared? What are the other bases for sharing, assistance, and friendship? What role do groups and coalitions play in shaping patterns of assistance? What is the role of aggression, retaliation, and status?

We expect that further evolutionary analyses will support the claim that different decision rules will be favored for long-term versus short-term relationships, for high- versus low-variance resources, and for other variations in conditions that we have not considered here. If this turns out to be true, then although some decision rules governing social exchange should be common across situations, others will differ in ways that are sensitively dependent on context. In other words, the social contract algorithms might contain different situation-specific modes of activation. For example, we expect the rules governing social exchange among close friends to differ somewhat from the rules governing social exchange among strangers.

Given an evolved architecture with a design of this kind, one can develop a coherent conceptual framework for understanding the ecological distribution of mental representations and expressed behaviors having to do with social exchange, both within and between groups (Sperber, 1985). Although the analysis of individual and cultural variation within the context of a universal human nature is too complex to review fully here, extended discussions can be found in Cosmides and Tooby (1987), Tooby and Cosmides (1989, 1990a), and Brown (1991). Briefly:

1. By virtue of being members of the human species, all humans are expected to have the same adaptive mechanisms, either at the level of developmental programs, which govern alternative developmental pathways, or (more likely in the case of social exchange) universally developing cognitive specializations.
2. In consequence, certain fundamental ways of thinking about social exchange will be the same everywhere, without needing to be socially transmitted.
3. As is standardly believed, social transmission may indeed shape social exchange. But it does so not by manufacturing the concept of social exchange de novo. Instead, there are probably certain specific avenues through which social transmission can act: for example, by influencing the valuation placed on items or actions, by providing information that helps one identify appropriate partners, or by providing information that allows one to identify appropriate contexts.
4. Our social exchange psychology should be highly context-dependent. Consequently, many dimensions along which social exchange varies, both within and between cultures, may be instances of evoked culture. The presence or magnitude of certain cues, conditions, or situations should cause mechanisms within our social exchange psychology to be calibrated; to be activated or inhibited; or to be switched into alternative modes of activation. When local circumstances trigger a particular mode of activation of the social contract algorithms, for example, this may cause a highly structured, multi-individual behavioral outcome. Therefore, when one sees similar patterns of social exchange in widely different parts of the world, one cannot assume that the similarity is determined primarily by social transmission; the similarity may instead by an instance of evoked culture.

One simple illustration of evoked culture involves the decision rules governing reciprocity in food sharing. As we describe below, a contextual variable—the presence or absence of "luck" in food acquisition—appears to activate different decision rules governing food sharing.

Luck and Sharing

One finding from the literature of evolutionary ecology on optimal foraging is that different kinds of sharing rules benefit the individual in different situations (see Kaplan & Hill, 1985). For example, when the variance in foraging success of the individual is greater than the variance for the band as a whole, band-wide food sharing buffers the variance. This can happen when one individual's success on a given day is unconnected to that of another.

Luck, skill, and effort all affect whether an individual finds food on a given day, but for certain food sources, luck is much more important than skill and effort. When success in finding food randomly varies a great deal from day to day, consider what would happen to a person who ate only that which he or she individually acquired. Some days that person would be starving; other days that person would have more to eat than he or she could possibly consume. It would be a feast or famine kind of life. Moreover, the temporary famines hit harder than the feasts can make up for. This is because (a) there is a zero point with food—death by starvation—and (b) the law of decreasing marginal utilities applies because we can only metabolize so much at one time—consequently, the fifth pound of meat eaten is far less valuable than the first pound. Under these circumstances, one is better off individually if one can redistribute food from periods of feast to periods of famine. There are two ways of doing this: through food storage or through pooling resources with others. Food storage is not an option for many hunter-gatherers, but pooling resources is: If two people average their returns, the variance decreases—one buys fewer periods of privation at the price of fewer periods of superabundance. By adding more individuals to the risk-pooling group, the variance may continue to decrease, making band-wide sharing an attractive system for hunter-gatherers facing certain conditions.

Thus, situations involving a random and frequent reversal of fortune can create substantial payoffs for cooperation. In effect, an individual can store food in the form of social obligations—by accepting food, others obligate themselves to reciprocate in the future. I may sacrifice by giving you some of my food today, but tomorrow I may be the one who is empty-handed and in need. For situations involving frequent, chance-driven reversals of fortune, the favored strategy involves sharing, from individuals who have food to those who do not. Luck plays an important role in hunting; consequently, hunter-gatherers frequently distribute game relatively equally to everyone in the band, no matter who found it or made a particular kill. Because it is a relatively high-variance activity, hunting may have been a particularly important driving force in the evolution of cognitive adaptations for social exchange (see Tooby & DeVore, 1987, for discussion).

By the same token, when variance in foraging success for an individual is low, the average long-term payoffs to sharing are less. If everyone reliably has access to the same goods, there is no particular benefit to sharing—one gains nothing by swapping the same goods at the same time. In this circumstance, an individual may be better off sharing just within his or her family, in accordance with kin selection, mating, and parenting principles.

Under low-variance conditions, not only might there be no benefit to sharing, there may be definite costs. When luck is eliminated as a factor, skill and effort remain. The smaller the role played by chance, the more differences between individuals in amount of food foraged will reflect differences in skill and effort. Under such circumstances, band-wide food sharing would simply redistribute food from those who expend more effort or are more skilled, to those who expend less effort or are less skilled. Sharing under these circumstances offers few—if any—intrinsic payoffs for those who have acquired more food. Without chance creating reversals of fortune, there is little reason to expect that the future will be different from the present and, therefore, little reason to expect that those with less food now will be in a better position to reciprocate in the future. Under these circumstances, then, one expects that (a) potential recipients will welcome sharing, but (b) potential donors will be more reluctant to share.

Consequently, the degree and source variance in resource acquisition were selection pressures that should have shaped the evolved architecture of our social exchange algorithms. Information about variance in foraging success should activate different modes of operation of these algorithms, with high variance due to chance triggering a psychology of sharing. To modern social scientists, factors such as variance in food acquisition may seem arcane and implausible because of their lack of connection to modern (middle-class) experience. But for our ancestors, food acquisition was a daily problem, as consequential as breathing. Daily decisions with respect to sharing had an unremitting impact on their lives and reproductive success, over hundreds of thousands of generations. In consequence, it is hard to see how our social psychology would not have been shaped by factors of this kind.

Obviously, this analysis of selection pressures is restricted to factors internal to foraging success. There are, of course, many other selection pressures that have shaped human social psychology over evolutionary time and hence many other factors that may lead to food sharing other than simple social exchange—kinship, love, parenting, sex, coercion, and status, for example. Moreover, even within the context of social exchange, the return on sharing food may be something other than food. Selection may have produced psychological mechanisms that cause highly productive foragers to share food without expecting any return of food, if, for example, by so doing others valued them highly and were therefore more disposed to render them aid when they were threatened, protect their children, grant them sexual access, and so on (Kaplan & Hill, 1985). Complicated though it may be, a more comprehensive understanding of social exchange eventually can be built up, element by element, by examining each selection pressure in turn and seeing whether our psychological mechanisms have the design features these selection pressures would lead one to expect.

In other words, the selection pressures analyzed in optimal foraging theory are one component of a task analysis, or, in David Marr's terms, a "computational theory," of the adaptive problem of foraging. It defines the nature of the problem to be solved and thereby specifies constraints that any mechanism that evolved to solve this problem can be expected to satisfy. In this case, optimal foraging theory suggests (a) that we should have content-specific information-processing mechanisms governing foraging and sharing, and (b) these mechanisms should be sensitive to information regarding variance in foraging success, causing us to prefer one set of sharing rules for high-variance items and another set for low-variance items.

The Ache: Within-Group Evidence for Evoked Culture

Kaplan and Hill's (1985) study of the Ache, a hunter-gatherer group living in eastern Paraguay, provides a particularly elegant test of the hypothesis just described, because it controls for "culture." Meat is a very high-variance food item among the Ache: On any given day, there is a 40% chance that a hunter will come back empty-handed (Kaplan, Hill, & Hurtado, 1990). Collected plant foods, in contrast, are very low-variance items. Kaplan and Hill found that the Ache engage in band-wide sharing of meat, whereas they share plant foods primarily within the nuclear family. Thus the same individuals, in the same culture, engage in different patterns of sharing for different foods, depending on the variance they experience in obtaining them.

The fact that meat is such a high-variance item also creates problems in cheater detection. If a man brings back no meat for seven days in a row, has he just had a run of bad luck, or has he been shirking? An Ache man's life and the life of his family depend on the long-term reciprocity relationships he has with the other hunters in his band. To accuse someone of cheating and ostracize him from the reciprocity network is a very serious matter. If the charge is false, then not only will the ostracized man's survival be jeopardized, but each member of the band will have lost a valuable reciprocation partner. If one is not sure, or if the suspected cheater is providing a net benefit even though it is less than he could provide if he tried harder, it might be better to continue the relationship.

The anthropologists who study the Ache know who the best hunters are, because they have recorded and weighed what each man brings back over long periods of time. Presumably, the Ache know also. But H. Kaplan (personal communication, 1991) reports that when he and his colleagues ask Ache men who the best hunters are, the question makes them very uncomfortable and they refuse to answer.[22] This is not due to a general cultural prohibition against accusing others of cheating. When the Ache are staying at a mission camp, acrimonious arguments erupt over whether various individuals are doing their fair share of work in the garden. Gardening, however, provides a low-variance source of food, making the punishment of cheaters less risky, and it occurs in a well-defined, observable location, making it easy to monitor who is, and who is not, cheating.

!Kung San Versus //Gana San: Between-Group Evidence for Evoked Culture

Resource variance can also explain differences between groups, evoking different cultures in response to different, variance-related local circumstances. For example, Cashdan (1980) found variance-related differences in sharing between groups of Kalahari San that mirror those found within Ache culture.

The Kalahari San are well known in anthropological circles for their economic and political egalitarianism. For example, the !Kung San, who experience extreme variability in the availability of food and water, have very strong social sanctions that reinforce sharing, discourage hoarding (calling someone "stingy" is a strong insult), and discourage displays of arrogance and authority. For example:

> The proper behavior of a !Kung hunter who has made a big kill is to speak of it in passing and in a deprecating manner . . . ; if an individual does not minimize or speak lightly of his own accomplishments, his friends and relatives will not hesitate to do it for him. (Cashdan, 1980, p. 116)

But it turns out that some San bands are more egalitarian than others, and their degree of egalitarianism is related to variance in their food supply. The //Gana San of the northeastern Kalahari are able to buffer themselves from variability in the food and water supply in ways that other San cannot, through a small amount of food cultivation (including a kind of melon that stores water in the desert environment) and some goat husbandry. In contrast to the !Kung, the //Gana manifest considerable economic inequality, they hoard more, they are more polygynous, and, although they have no clear-cut authority structure, wealthy, high-status //Gana men are quick to claim that they speak for others and that they are the "headman"—behavior that would be considered unconscionable among the !Kung. Again, even though the !Kung and the //Gana are culturally similar in many ways—they share the same encompassing "meme-pool," so to speak—their social rules regarding sharing and economic equality differ, and these differences track the variance in their food and water supplies.

Local Conditions and Evoked Culture

It is difficult to explain these phenomena simply as the result of cultural transmission, at least in any traditional sense. Among the Ache of Paraguay, the same individuals share food types with different variances differently. Half way around the world, in Africa, two different groups of Kalahari San manifest what appear to be the same differential sharing patterns in response to the same variable—variance. A parsimonious explanation is that these social norms and the highly patterned behaviors they give rise to are evoked by the same variable.

Because foraging and sharing are complex adaptive problems with a long evolutionary history, it is difficult to see how humans could have escaped evolving highly structured domain-specific psychological mechanisms that are well designed for solving them. These mechanisms should be sensitive to local informational input, such as information regarding variance in the food supply. This input can act as a switch, turning on and off different modes of activation of the appropriate domain-specific mechanisms. The experience of high variance in foraging success should activate rules of inference, memory retrieval cues, attentional mechanisms, and motivational mechanisms. These should not only allow band-wide sharing to occur, but should make it seem fair and appealing. The experience of low variance in foraging success should activate rules of inference, memory retrieval cues, attentional mechanisms, and motivational mechanisms that make within-family sharing possible and appealing, but that make band-wide sharing seem unattractive and unjust. These alternative modes of activation of the domain-specific mechanisms provide the core knowledge that must be mutually manifest (see Sperber & Wilson, 1986) to the various actors for band-wide or within-family sharing to occur. This core knowledge can then organize and provide points of attachment for symbolic activities that arise in these domains.

If this notion of evoked culture is correct, then one should not expect cultural variation to vary continuously along all imaginable dimensions. The free play of human creativity may assign relatively arbitrary construals to elements in some areas of life, such as the number of gods or the appropriate decoration on men's clothing. But in other areas of life one might expect there to be a limited number of recurring patterns, both within and across cultures. For certain domains of human activity, people from very different places and times may "reinvent" the same kinds of concepts, valuations, social rules, and customs (see Tooby & Cosmides, this volume). In short, such alternative modes of activation in psychological mechanisms can create alternative sets of complexly patterned social rules and activities. These will emerge independently, that is, in the absence of direct cultural transmission, in culture after culture, when the individual members are exposed to the informational cues that activate these alternative modes.

Cross-cultural studies of social exchange by Fiske provide support for this notion (Fiske, 1990, 1991a). Based on his field studies of the Moose ("Mossi") of Burkina Faso and his review of the anthropological literature, Fiske argues that the human mind contains four alternative implicit models of how sharing should be conducted, which are used to generate and evaluate social relations. These models are implicit in the sense that they are acted on unreflectively and without conscious awareness; indeed, they may never have been explicitly stated by any member of the culture. Nevertheless, "these shared but unanalyzed, tacit models for Moose social relations allow them to generate coordinated, consistent, and culturally comprehensible interactions of four contrasting types" (Fiske, 1990, pp. 180–181). For example, one of Fiske's four models is communal sharing of the kind used by the Ache in distributing hunted meat; another is "market pricing"—the kind of explicit contingent exchange that occurs when two people explicitly agree to trade, say, honey for meat or money for milk.

Varieties of Hunter-Gatherer Exchange

Whether or not Fiske's specific taxonomy of four categories is exactly the correct way to capture and characterize the limited set of modes whereby humans engage in social exchange, we very much agree with this general framework for conceptualizing cultural variation in social exchange. If human thought falls into recurrent patterns from place to place and from time to time, this is because it is the expression of, and anchored in, universal psychological mechanisms. If there is a limited set of such patterns, it is because different modes of activation of the algorithms regulating social exchange solved different adaptive problems that hunter-gatherers routinely faced. Consequently, clues as to how many modes of activation the social contract algorithms have, what the structure of each mode might be, and what kinds of circumstances can be expected to activate each mode can be found by investigating the various forms of social exchange that hunter-gatherers engage in, as well as the conditions under which each form of exchange arises.

Despite the common characterization of hunter-gatherer life as an orgy of indiscriminate, egalitarian cooperation and sharing—a kind of retro-utopia—the archaeological and ethnographic record shows that hunter-gatherers engaged in a number of different forms of social exchange (for an excellent review of hunter-gatherer economics, see Cashdan, 1989). Communal sharing does not exhaust the full range of exchange in such societies. Hunter-gatherers also engage in explicit contingent exchange—Fiske's "market pricing"—in which tools and other durable goods are traded between bands, often in networks that extend over vast areas. A common form of trade is formal gift exchanges with carefully chosen partners from other bands. For instance, aboriginal Australians traded tools such as sting ray spears and stone axes through gift exchanges with partners from neighboring bands. These partnerships were linked in a chain that extended 620 km, from the coast, where sting ray spears were produced, to the interior, where there were quarries where the stone axes could be produced. Here, environmental variation in the source of raw materials for tool making allowed gains from trade based on economic specialization, and the laws of supply and demand seemed to operate: At the coast, where sting ray spears were common, it took more of them to buy an ax than in the interior, where spears were dear and axes cheap

Sharp, 1952). Similarly, the !Kung of the Kalahari desert engage in a system of lelayed reciprocal gift giving called "hxaro" (Weissner, 1982; Cashdan, 1989), hrough which they trade durable goods such as blankets and necklaces.

Unpredictable variation in rainfall and game makes access to land and water esources another important "item of trade" between hunter-gatherer bands and creites situations in which a kind of implicit one-for-one reciprocity prevails (Fiske's 'equality matching"). For instance, a !Kung band that is caught in a drought will "visit relatives" in a band that is camped in an area that is experiencing more rainfall (Cashlay, 1989). Indeed, hxaro partners are chosen carefully, not only for their ability to confer durable goods, but also to provide alternative residences in distant places during times of local scarcity (Weissner, 1982). And before using another band's water hole or land, the !Kung are expected to ask permission; reciprocity in access to water holes is extremely important to the !Kung, who live in a desert with few permanent sources of water. Although formal permission is almost always granted, as the implicit rules of one-for-one reciprocity require, if the hosts really don't want to accommodate their guests, they make them feel unwelcome, thereby subtly encouraging them to leave (Cashdan, 1989).

Although authoritarian social relations are unusual among the few remaining modern hunter-gatherer groups, this is probably a by-product of their having been pushed into marginal environments by the peoples of agricultural and industrial cultures. Variance in the food supply is high in harsh environments like the Kalahari desert, and band-wide communal sharing is advantageous for high-variance resources. But as variance is buffered, as in the //Gana San example discussed earlier, more inequality and more authority-ranking relationships develop. This process was, for example, quite pronounced in the hunter-gatherer societies of the Pacific Northwest. The Pacific Northwest was so rich in fish and game that the hunter-gatherers living there could afford to be relatively sedentary. These people developed stable, complex societies that were so hierarchical that some of them even included a slave class formed from prisoners of war (Drucker, 1983; Donald, 1983). Of course, the distribution of goods and services that occurs between individuals of different rank is often determined by an uneasy mixture of coercion, threat, and exchange.

This is not the place to attempt a full computational theory of the various modes of activation of the social contract algorithms. But even these brief examples drawn from hunter-gatherer life provide some hints as to what might be relevant variables in such an analysis: variance in the food supply; degree of kinship; status or rank; whether a relationship is long- or short-term; whether one is in daily contact (communal sharing; implicit deferred reciprocity) or only rare contact (explicit contingent exchange); whether storage is possible; whether the group is sedentary enough for inequalities in wealth to accumulate; whether gaining a resource requires close, interdependent cooperation; whether people are trading different resources or dividing the same resource; whether an external, consensual definition of "equal portion" is feasible; whether an individual can control access to a resource, and thereby "own" it; and so on (see also McGrew & Feistner, this volume).

To understand social exchange in all its various forms, the adaptive problems that selected for different decision rules must be precisely defined in the form of computational theories. The computational theories can then be used to generate hypotheses about the design features that characterize the different modes of activation of the social contract algorithms. Psychological experiments of the kind described earlier in this chapter would allow one to test among these hypotheses and thereby develop a detailed map of the situation-specific cognitive processes that create these different modes of activation. Once we know what situational cues activate each set of decision rules, we should be able to predict a great deal of cultural variation.

Interpreting Other Cultures and Understanding Cultural Change

Significant aspects of cultural variation in social exchange can be readily reconciled with a universal human nature through applying the concept of evoked culture. The various sets of decision rules governing social exchange will be universal, but which sets are activated will differ from situation to situation within a culture, as well as between cultures. For example, in American middle-class culture different exchange rules apply to different aspects of a dinner party (Fiske, 1991b). Invitations are sometimes governed by one-for-one reciprocity—an implicit rule such as "If you had me to your home for dinner, then at some point I must invite you to dinner." But food sharing at the party is governed by the same kind of communal sharing rules that characterize Ache meat sharing. Obtaining the food that is served is governed by explicit contingent exchange at a grocery store, and seating at the dinner table is sometimes determined by rank or status (as for example, at diplomatic dinners, birthday parties, or in certain traditional families).

The point is that communal sharing, explicit contingent exchange, equality matching, and so on, are not unique to American culture: The same sets of decision rules appear in other cultures as well, but local circumstances cause them to be applied to different situations (Fiske, 1990, 1991a). Whereas all food at an American dinner party is shared communally, this is not true on Ache foraging trips: Meat is shared communally at the level of the entire band, but plant foods are not. In many cultures, men engage in explicit contingent exchange to procure wives: One man will buy another man's daughter (see Wilson & Daly, this volume). In other cultures, men do not buy wives, but instead can engage in explicit contingent exchange with a woman to gain temporary sexual access to her. In still other cultures, the use of explicit contingent exchange is illegal in both circumstances (but may still be understood and occasionally practiced).

Fiske argues that in relatively stable, traditional societies there is a tacit consensus about which decision rules to apply in which situation. To apply the wrong decision rules to a situation can be uncomfortable, insulting, or even shocking: At the end of an American dinner party, one does not pull out a wallet and offer to pay the hosts for their services. Similarly, when Americans are sitting with friends or co-workers, they might spontaneously offer to split a sandwich, but they almost never spontaneously pull out their wallets and offer money. Indeed, figuring out which decision rules a culture applies to which situations is part of what it means to understand another culture (Fiske, 1990). On this view, "interpreting another culture" is not usually a matter of absorbing wholly new systems of culturally alien semantic relations. Instead, interpreting another culture is a matter of learning how the evolved set of meanings that we have come to assign to one set of objects or elements in a situation are, in another culture, assigned to a different set.

New events of all kinds, from migrations to natural disasters to new technologies, create culturally unprecedented circumstances in which there is no within-culture consensus about which exchange rules are appropriate. In the United States, for example, there is a vigorous debate over which form of exchange should apply when a woman wants to be a surrogate mother for an infertile couple. Many women prefer explicit contingent exchange in which they are paid money for their labor (so to speak). But other Americans argue that surrogacy should occur—if at all—only among close friends and relatives who participate in informal communal sharing relationships and that women should be legally prohibited from granting access to their wombs on the basis of explicit contingent exchange.

Where do the impulses—or, more accurately—the decision rules come from that lead individuals or entire cultures to reject an existing practice or to invent or adopt something new? Transmission models can account for stable transmission of existing attitudes and cultural forms but intrinsically have no way to account for cultural change, or indeed any nonimitated individual act. The existence of a species-typical evolved psychology fills in this missing gap. It provides a basis from which one can interpret individual action, minority dissent, and the emergence of a new consensus. Dramatic new circumstances may evoke new attitudes overnight, as when the Battle of Britain changed the attitudes and sharing practices of Londoners, or when depictions of earthquakes and other natural disasters prompt people to donate food and other assistance. Even where one is dealing with the spread of new cultural forms through transmission, however, the dynamics are powerfully structured by our content-sensitive evolved psychology (for a lucid discussion of the "epidemiology" of beliefs and other representations, see Sperber, 1985, 1990).

Consider the political and moral debate concerning the homeless in the United States. Those with opposing postures concerning how much to help the homeless frame their positions in ways that exploit the structure of this evolved psychology. One persistent theme among those who wish to motivate more sharing is the idea of "there but for fortune, go you or I." That is, they emphasize the random, variance-driven dimensions of the situation. The potential recipient of aid is viewed as worthy because he or she is the unlucky victim of circumstances, such as unemployment, discrimination, or mental illness. On the other hand, those who oppose an increase in sharing with the homeless emphasize the putatively chosen or self-caused dimensions of the situation. Potential recipients are viewed as unworthy of aid because they "brought it on themselves": They are portrayed as able-bodied but lazy, or as having debilitated themselves through choosing to use alcohol and other drugs. The counterresponse from those who want to motivate more sharing is to portray drug use not as a choice, but as a sickness, and so on.

If cultural meanings were truly arbitrary, then, cross-culturally, donors would be just as likely to view people as "worthy of assistance" when they have "brought it on themselves" as when they have been "the victims of bad luck." Indeed, if all were arbitrary, somewhere one should find a culture in which potential donors are most eager to help those who are *more* fortunate than themselves, merely *because* the potential recipients are more fortunate (and not, say, because they hope for something in return).

Finally, although our cognitive mechanisms evolved to promote adaptive decisions in the Pleistocene, they do not necessarily produce adaptive decisions under evolutionarily novel modern circumstances (see Symons, this volume). For example, if individual variance in obtaining alcohol is greater than group variance for homeless alcoholics who camp out in the same alley, this circumstance might activate decision rules that promote communal sharing of alcohol, even though these people's mutual generosity would be slowly killing them.

CONCLUSIONS

Human reason has long been believed to be the paradigm case of the impartial, content-blind, general-purpose process. Further, it has been viewed as the faculty that distinguished humans from all other animals, and the very antithesis of "instinct." But if even reasoning turns out to be the product of a collection of functionally specialized, evolved mechanisms, most of which are content-dependent and content-imparting, then this has implications far beyond the study of reasoning. The presumption that psychological mechanisms are characteristically general-purpose and content-free would no longer be tenable: Such hypotheses should no longer be accorded the privileged status and the near-immunity from question that they have customarily received. Instead, domain-general and domain-specific hypotheses should be given equal footing and evaluated solely on their ability to be turned into genuine, well-specified models that actually account for observed phenomena. Guided by such tenets, we may discover that the human mind is structurally far richer than we have suspected and contains a large population of different mechanisms.

We have used as a test case the intersection of reasoning and social exchange. The results of the experiments discussed herein directly contradict the traditional view; they indicate that the algorithms and representations whereby people reason about social exchange are specialized and domain-specific. Indeed, there has been an accumulation of "evidence of special design" (Williams, 1966), indicating the presence of an adaptation. The results are most parsimoniously explained by positing the existence of "specialized problem-solving machinery" (Williams, 1966)—such as cost-benefit representations and cheater detection procedures—that are well designed for solving adaptive problems particular to social exchange. Moreover, they cannot be explained as the by-product of mechanisms designed for reasoning about classes of problems that are more general than social contracts, such as "all propositions," or even the relatively restricted class of "all permissions." In addition, the pattern of results indicate that this specialized problem-solving machinery was not *built* by an evolved architecture that is general-purpose and content-free (see Implications for Culture, this chapter, and Cosmides, 1989). In other words, the empirical record is most parsimoniously explained by the hypothesis that the evolved architecture of the human mind contains functionally specialized, content-dependent cognitive adaptations for social exchange. Such mechanisms, if they exist, would impose a distinct social contract conceptual organization on certain social situations and impart certain meanings to human psychological, social, and cultural life. We suggest these evolved algorithms constitute one functional subunit, out of many others, that are linked together to form a larger faculty of social cognition (e.g., Jackendoff, 1991).

The results of the experiments discussed herein undermine two central tenets of the Standard Social Science Model (Tooby & Cosmides, this volume). First, they undermine the proposition that the evolved architecture of the human mind contains a single "reasoning faculty" that is function-general and content-free. Instead, they support the contrary contention that human reasoning is governed by a diverse collection of evolved mechanisms, many of which are functionally specialized, domain-specific, content-imbued, and content-imparting (see Tooby & Cosmides, this volume). According to this contrary view, situations involving threat, social exchange, hazard, rigid-object mechanics, contagion, and so on each activate different sets of functionally specialized procedures that exploit the recurrent properties of the corresponding domain in a way that would have produced an efficacious solution under Pleistocene conditions. On this view, the human mind would more closely resemble an intricate network of functionally dedicated computers than a single general-purpose computer. The second tenet that these results undermine is the proposition that all contentful features of the human mind are "socially constructed" or environmentally derived. In its place, this research supports the view that the human mind imposes contentful structure on the social world, derived from specialized functional design inherent in its evolved cognitive architecture.

The conceptual integration of evolutionary biology with cognitive psychology offers something far more valuable than general arguments. The analysis of the computational requirements of specific adaptive problems provides a principled way of identifying likely new modules, mental organs, or cognitive adaptations, and thereby opens the way for extensive empirical progress. By understanding these requirements, one can make educated guesses about the design features of the information-processing mechanisms that evolved to solve them. Turning knowledge of the adaptive problems our ancestors faced over evolutionary time into well-specified computational theories can therefore be a powerful engine of discovery, allowing one to construct experiments that can capture, document, and catalog the functionally specialized information-processing mechanisms that collectively constitute much (or all) of our "central processes." In effect, knowledge of the adaptive problems humans faced, described in explicitly computational terms, can function as a kind of Rosetta Stone: It allows the bewildering array of content effects that cognitive psychologists routinely encounter—and usually disregard—to be translated into meaningful statements about the structure of the mind. The resulting maps of domain-specific information-processing mechanisms can supply the currently missing accounts of how the human mind generates and engages the rich content of human culture, behavior, and social life.

ACKNOWLEDGMENTS

We thank Jerry Barkow, David Buss, Martin Daly, Lorraine Daston, Gerd Gigerenzer, Steve Pinker, Paul Romer, Roger Shepard, Don Symons, Phil Tetlock, Dan Sperber, Valerie Stone, and Margo Wilson for productive discussions of the issues addressed in this chapter or for their useful comments on previous drafts. This chapter was prepared, in part, while the authors were Fellows at the Center for Advanced Study in the Behavioral Sciences. We are grateful for the Center's support, as well as that provided by the Gordon P. Getty Trust, the Harry Frank Guggenheim Foundation, NSF Grant BNS87-00864 to the Center, NSF Grant BNS85-11685 to Roger Shepard, and NSF Grant BNS91-57449 to John Tooby.

NOTES

1. For example, if C_i and B_i refer to decreases and increases in *i*'s reproduction, then a decision rule that causes *i* to perform act *Z* if, and only if, C_i of doing $Z < 0$ would promote its own

Darwinian fitness, but not its inclusive fitness. In contrast, a decision rule that causes i to perform act *Z* if, and only if, (C_i of doing *Z*) < (B_j of i's doing *Z*) × r_{ij} would promote its inclusive fitness, sometimes at the expense of its Darwinian fitness. The first decision rule would be at a selective disadvantage compared with the second one, because it can make copies of itself only through its bearer, and not through its bearer's relatives. For this reason, designs that promote their own inclusive fitness tend to replace alternative designs that promote Darwinian fitness at the expense of inclusive fitness.

Although this example involves helping behavior, kin selection theory applies to the evolution of nonbehavioral design features as well, for example, to the evolution of aposematic coloration of butterfly wings. In principle, one can compute the extent to which a new wing color affects the reproduction of its bearer and its bearer's kin, just as one can compute the extent to which an *action* affects the reproduction of these individuals.

2. Other models of social exchange are possible, but they will not change the basic conclusion of this section: that reciprocation is necessary for the evolution of social exchange. For example, the Prisoner's Dilemma assumes that enforceable threats and enforceable contracts are impossibilities (Axelrod, 1984), assumptions that are frequently violated in nature. The introduction of these factors would not obviate reciprocation—in fact, they would enforce it.

3. Following Marr, 1982, we would like to distinguish between the cognitive program itself and an abstract characterization of the decision rule it embodies. Algorithms that differ somewhat in the way they process information may nevertheless embody the same decision rule. For example, the algorithms for adding Arabic numerals differ from those for adding Roman numerals, yet they both embody the same rules for addition (e.g., that $A + B = B + A$) and therefore yield the same answer (Marr, 1982).

4. These selection pressures exist *even in the absence of competition for scarce resources.* They are a consequence of the game-theoretic structure of the social interaction.

5. The game "unravels" if they do. If we both know we are playing three games, then we both know we will mutually defect on the last game. In practice, then, our second game is our last game. But we know that we will, therefore, mutually defect on that game, so, in practice, we are playing only one game. The argument is general to any known, fixed number of games (Luce & Raiffa, 1957).

6. The cost-benefit values that these algorithms assign to items of exchange should be correlated with costs and benefits to fitness in the environment in which the algorithms evolved; otherwise, the algorithms could not have been selected for. But these assigned values will not necessarily correlate with fitness in the modern world. For example, our taste mechanisms assess fat content in food and our cognitive system uses this cue to assign food value: We tend to like food "rich" (!) in fat, such as ice cream, cheese, and marbled meat. The use of this cue is correlated with fitness in a hunter-gatherer ecology, where dietary fat is hard to come by (wild game is low in fat). But in modern industrial societies, fat is cheap and plentiful, and our love of it has become a liability. The environment changed in a way that lowered the cue validity of fat for fitness. But our cognitive system, which evolved in a foraging ecology, still uses it as a cue for assigning food value.

Given the long human generation time, and the fact that agriculture represents less than 1% of the evolutionary history of the genus *Homo*, it is unlikely that we have evolved any complex adaptations to an agricultural (or industrial) way of life (Tooby & Cosmides, 1990a). Our ancestors spent most of the last 2 million years as hunter-gatherers, and our primate ancestors before the appearance of *Homo* were foragers as well, of course. The very first appearance of agriculture was only 10,000 years ago, and it wasn't until about 5,000 years ago that a significant fraction of the human population was engaged in agriculture.

7. Interpreting the statement as a biconditional, rather than as a material conditional, will also lead to error. Consider a situation in which you gave me your watch, but you did not take my $20. This would have to be considered a violation of the rule on a biconditional interpretation, but it is not necessarily cheating. If I had not offered you the $20, then I would have cheated. But if I had offered it and you had refused to take it, then no cheating would have occurred on either of our parts. In this situation, your behavior could be characterized as altruistic or foolish, but not as cheating. Distinctions based on notions such as "offering" or intentionality are not part of the definition of a violation in the propositional calculus.

8. Indeed, one expects learning under certain circumstances: The genome can "store" information in the environment if that information is stably present (Tooby & Cosmides, 1990a).

9. Of course, if two developmental trajectories have different reproductive consequences, one will be favored over the other.

10. The drinking-age problem is also a social contract from the point of view of those who enacted the law. Satisfying the age requirement before drinking beer provides those who enacted the law with a benefit: People feel that the roads are safer when immature people are not allowed to drink. Although satisfying the requirement in a social contract will often cause one to incur a cost, it need not do so (see Cosmides & Tooby, 1989). The requirement is imposed not because it inflicts a cost on the person who must satisfy it, but because it creates a situation that benefits the recipient, which the recipient believes would not occur if the requirement were not imposed.

Consider the following social contract: "If you are a relative of Nisa's, then you may drink from my water hole." A hunter-gatherer may make this rule because she wants to be able to call on Nisa for a favor in the future. A given person either is, or is not, Nisa's relative; it would therefore be odd to say that being Nisa's relative inflicts a cost on one. Nevertheless, it is the requirement that must be satisfied to gain access to a benefit, and it was imposed because it creates a situation that can benefit the person who imposed it. This is why Cheng and Holyoak's (1989) distinction between "true" social exchange, where the parties incur costs, and "pseudo" social exchange, where at least one party must meet a requirement that may not be costly, constitutes a misunderstanding of social contract theory and the basic evolutionary biology that underlies it. Social exchange is the reciprocal provisioning of benefits, and the fact that the delivery of a benefit may prove costly is purely a by-product.

11. So far, the evidence suggests that we also have specialized procedures for reasoning about threats and precautions, for example.

12. No criticism of the experimenters is implied; these experiments were not designed for the purpose of testing social contract theory.

13. What if people read in a "may" that refers to obligation, rather than to possibility? That is, after all, a prediction of social contract theory. Logicians have tried to create "deontic logics": rules of inference that apply to situations of obligation and entitlement. Social contract theory is, in fact, a circumscribed form of deontic logic. But could subjects be using a generalized form of deontic logic? Manktelow and Over (1987) say that the answer is not clear because deontic logicians do not yet agree: According to some, no cards should be chosen on the switched social contracts; according to others, *not-P & Q* should be chcsen. Because the rules of inference in social contract theory include the concepts of entitlement and obligation, it can be thought of as a specialized, domain-specific deontic logic. But we doubt that people have a generalized deontic logic. If they did, then non-social contract problems that involve obligation should elicit equally high levels of performance. But this is not the case, as will be discussed later in the chapter.

14. Even if this hypothesis were true, one would still have to explain *why* social contract problems are easier to understand, or more interesting, than other situations. After all, there is nothing particularly complicated about the situation described in a rule such as "If a person eats red meat, then that person drinks red wine." Social contract problems could be easier to understand, or more interesting, precisely because we do have social contract algorithms that organize our experience in such situations. Consequently, showing that social contract problems afford clear thinking about a wide variety of problems would not eliminate the possibility that there are social contract algorithms; it would simply cast doubt on the more specific claim that this set of algorithms includes a procedure specialized for cheater detection.

15. We would like to point out that the relationship between psychology and evolutionary biology can be a two-way street. For example, one could imagine models for the emergence of

stable cooperation that require the evolution of a mechanism for altruism detection. If the selection pressures required by these models were present during hominid evolution, they should have left their mark on the design of our social contract algorithms. Finding that people are not good at detecting altruists casts doubt on this possibility, suggesting altruists were too rare to be worth evolving specialized mechanisms to detect, and hence gives insight into the kind of selection pressures that shaped the hominid line.

16. For example, instead of asking subjects to "indicate only those card(s) you definitely need to turn over to see if any of these boys have broken the law," the altruist version asked them to "indicate only those card(s) you definitely need to turn over to see if any of these boys have behaved altruistically with respect to this law."

17. Indeed, on the altruist detection problems in which the rule was a social law, more subjects detected *cheaters* than detected altruists! (This result was 64% in the altruist version; 44% in the selfless version.) It is almost as if, when it comes to a social law, subjects equate altruistic behavior with honorable behavior—i.e., with the absence of cheating. (This may be because for many social laws, such as the drinking age law, "society"—i.e., the individuals who enacted the law—benefits from the total configuration of events that ensues when the law is obeyed.) This was not true of the personal exchange laws, where it is easy to see how the other party benefits by your paying the cost to them but not accepting the benefit they have offered in return. (For the private exchange problems, only 16% of subjects chose the "look for cheaters" answer in the two altruist versions; 8% and 4%, in the selfless versions.)

18. Manktelow and Over (1987) point out that people do understand what conditions constitute a violation of a conditional rule, even when it is an abstract one. Hence the failure to perform well on the no cheating version cannot be attributed to subjects' not knowing what counts as a violation. (This fact may seem puzzling at first. But one can know what counts as a violation without being able to use that knowledge to generate falsifying inferences, as the failure to choose *P & not-Q* on abstract Wason selection tasks shows. Two separate kinds of cognitive processes appear to be involved. An analogy might be the ease with which one can recognize a name that one has been having trouble recalling.)

19. It is difficult to tell a permission from an obligation because both involve obligation and because there are no criteria for distinguishing the two representational formats ("If action is taken, then precondition must be satisfied" versus "If condition occurs, then action must be taken"). "Conditions" and "preconditions" can, after all, be "actions." The primary difference seems to be a time relation: If the obligation must be fulfilled before the action is taken, it is a permission. If the obligation can be fulfilled after a condition (which can be an "action taken") occurs, then it is an obligation. A social contract of the form, "If you take the benefit, then you must pay the cost" would be considered a permission if you were required to pay the cost before taking the benefit, but an obligation if you had first taken the benefit, thereby incurring the obligation to pay the cost.

20. To choose *not-P & Q*, one would have to interpret "If an employee gets a pension, then that employee must have worked for the firm for at least 10 years" as also implying "If an employee has worked for the firm for at least 10 years, then that employee *must* be given a pension." Social contract theory predicts that the one statement will be interpreted as implying the other, but permission schema theory does not. In fact, its translation rules (the four production rules) bar this interpretation. The rule presented to subjects—"If an employee gets a pension, then that employee must have worked for the firm for at least 10 years"—has the linguistic format of rule 1 of the permission schema—"If the action is to be taken, then the precondition must be satisfied." Rule 1 can be taken to imply rules 2, 3, and 4, but not other rules. By rule 3, the rule stated in the problem would translate to "If an employee has worked for the firm for at least 10 years, then that employee *may* be given a pension"—not that the employee *must* be given a pension.

21. Or equally low performance. Cheng & Holyoak have provided very little theory concerning what elements in a situation can be expected to activate the permission schema. Although they have suggested that the provision of a rationale or social purpose helps, they have never defined what counts as such, and there are (social contract) permission rules that lack rationales that nevertheless produce the effect (Cosmides, 1989). The problems that we tested here clearly stated that the rule is a law made by authorities, which ought to clarify that they are permission rules and prevent subjects from interpreting them as descriptive rules. If this is sufficient to activate a permission schema, then performance on all three problems should be equally high. But none of the problems contains or suggests a rationale. So if one were to claim that rationales are necessary, then performance on all three problems should be equally low. Either way, performance should not vary across the three problems.

22. This is the kind of situation that Nesse and Lloyd (this volume) suggest might call for benevolent self-deception. Although one memory module may be keeping an account of the other person's failure to contribute his fair share, this information might not be fed into the mechanisms that would cause an angry reaction to cheating. By preventing an angry reaction, this temporary encapsulation of the information would permit one to continue to cooperate with the suspected cheater. This situation would continue as long as one is still receiving a net benefit from the other person, or until it becomes sufficiently clear that the other person is cheating rather than experiencing a run of bad luck. At that point, the accounts kept by the one module would be fed into other modules, provoking an angry, recrimination-filled reaction.

REFERENCES

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211,* 1390–1396.

Axelrod, R. (1984). *The evolution of cooperation.* New York: Basic Books.

Boyd, R. (1988). Is the repeated prisoner's dilemma a good model of reciprocal altruism? *Ethology and Sociobiology, 9,* 211–222.

Brown, D. (1991). *Human universals.* New York: McGraw-Hill.

Cashdan, E. (1980). Egalitarianism among hunters and gatherers. *American Anthropologist, 82,* 116–120.

Cashdan, E. (1989). Hunters and gatherers: Economic behavior in bands. In S. Plattner (Ed.), *Economic Anthropology* (pp. 21–48). Stanford: Stanford University Press.

Cheng, P., & Holyoak, K. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17,* 391–416.

Cheng, P., & Holyoak, K. (1989). On the natural selection of reasoning theories. *Cognition, 33,* 285–313.

Cheng, P., Holyoak, K., Nisbett, R., & Oliver, L. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18,* 293–328.

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. Dupré (Ed.), *The latest on the best: Essays on evolution and optimality* (p. 277–306). 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. (in prep., a). The logic of threat: Evidence for another cognitive adaptation?

Cosmides, L., & Tooby, J. (in prep., b). Social contracts, precaution rules, and threats: How to tell one schema from another.

Cosmides, L., & Tooby, J. (in prep., c). Is the Wason selection task an assay for production rules in mentalese?

Dawkins, R. (1982). *The extended phenotype.* San Francisco: Freeman.

Dawkins, R. (1986). *The blind watchmaker.* New York: Norton.

Donald, L. (1983). Was Nuu-chah-nulth-aht (Nootka) society based on slave labor? Ecology and political organization in the Northwest coast of America. In E. Tooker (Ed.), *The development of political organization in native North America* (pp. 108–119). Washington, DC: American Ethnological Society.

Drucker, P. (1983). Ecology and political organization in the Northwest coast of America. In E. Tooker (Ed.), *The development of political organization in native North America* (pp. 86–96). Washington, DC: American Ethnological Society.

Evans, J.St.B.T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75,* 457–468.

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. (1990). Relativity within Moose ("Mossi") culture: Four incommensurable models for social relationships. *Ethos, 18,* 180–204.

Fiske, A. P. (1991a). *Structures of social life: The four elementary forms of human relations.* New York: Free Press.

Fiske, A. P. (1991b, June). *Innate hypotheses and cultural parameters for social relations.* Paper presented at the Society for Philosophy and Psychology, San Francisco, CA.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Frumhoff, P. C., & Baker, J. (1988). A genetic component to division of labour within honey bee colonies. *Nature, 333,* 358–361.

Frumhoff, P. C., & Schneider, S. (1987). The social consequences of honey bee polyandry: The effects of kinship on worker interactions within colonies. *Animal Behaviour, 35,* 255–262.

Gigerenzer, G., & Hug, K. (in press). Reasoning about social contracts: Cheating and perspective change. Institut für Psychologie, Universität Salzburg, Austria.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73,* 407–420.

Hanken, J., & Sherman, P. W. (1981). Multiple paternity in Belding's ground squirrel litters. *Science, 212,* 351–353.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. Parts I, II. *Journal of Theoretical Biology, 7,* 1–52.

Holmes, W. G., & Sherman, P. W. (1982). The ontogeny of kin recognition in two species of ground squirrels. *American Zoologist, 22,* 491–517.

Isaac, G. (1978). The food-sharing behavior of protohuman hominids. *Scientific American, 238,* 90–108.

Jackendoff, R. (1991, June). Is there a faculty of social cognition? Presidential address to the Society for Philosophy and Psychology, San Francisco, CA.

Johnson-Laird, P. N. (1982). Thinking as a skill. *Quarterly Journal of Experimental Psychology, 34A,* 1–29.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness.* Cambridge, MA: Harvard University Press.

Kaplan, H., & Hill, K. (1985). Food sharing among Ache foragers: Tests of explanatory hypotheses. *Current Anthropology, 26,* 223–239.

Kaplan, H., Hill, K., & Hurtado, A. M. (1990). Risk, foraging and food sharing among the Ache. In E. Cashdan (Ed.), *Risk and uncertainty in tribal and peasant economies.* Boulder: Westview Press.

Lee, R., & DeVore, I. (Eds.). (1968). *Man the hunter.* Chicago: Aldine.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey.* New York: Wiley.

Malinowski, B. (1922). *Argonauts of the Western Pacific.* New York: Dutton Press.

Maljković, V. (1987). *Reasoning in evolutionarily important domains and schizophrenia: Dissociation between content-dependent and content-independent reasoning.* Unpublished undergraduate honors thesis, Department of Psychology, Harvard University.

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. (1987). Reasoning and rationality. *Mind & Language, 2,* 199–219.

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). London: Wiley.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. San Francisco: Freeman.

Marr, D., & Nishihara, H. K. (1978, October). Visual information processing: Artificial intelligence and the sensorium of sight. *Technology Review,* 28–49.

Mauss, M. (1925/1967). *The gift: Forms and functions of exchange in archaic societies.* (I. Cunnision, Trans.). New York: Norton.

Maynard Smith, J. (1964). Group selection and kin selection. *Nature, 201,* 1145–1147.

Maynard Smith, J. (1982). *Evolution and the theory of games.* Cambridge: Cambridge University Press.

Packer, C. (1977). Reciprocal altruism in *Papio annubis. Nature, 265,* 441–443.

Pollard, P. (1982). Human reasoning: Some possible effects of availability. *Cognition, 10,* 65–96.

Quine, W.V.O. (1969). *Ontological relativity and other essays.* New York: Columbia University Press.

Sharp, L. (1952). Steel axes for stone age Australians. *Human Organization, 11,* 17–22.

Sherman, P. W. (1977). Nepotism and the evolution of alarm calls. *Science, 197,* 1246–1253.

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

Sperber, D. (1975). *Rethinking Symbolism* (A. Morton, Trans.). Cambridge: 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.*

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition.* Oxford: Blackwell.

Stone, V., & Cosmides, L. (in prep.). Do people have causal reasoning schemas?

Thornhill, R. (1991). The study of adaptation. In M. Bekoff and D. Jamieson (Eds.), *Interpretation and explanation in the study of behavior.* Boulder: West View Press.

Tooby, J. (1985). The emergence of evolutionary psychology. In D. Pines (Ed.), *Emerging Syntheses in Science. Proceedings of the Founding Workshops of the Santa Fe Institute.* Santa Fe, NM: The Santa Fe Institute.

Tooby, J., & Cosmides, L. (1989). Evolutionary psychology and the generation of culture, part I. Theoretical considerations. *Ethology & Sociobiology, 10,* 29–49.

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

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.

Trivers, R. L., & Hare, H. (1976). Haplodiploidy and the evolution of social insects. *Science, 191,* 249-263.

de Waal, F.B.M. (1982). *Chimpanzee politics: Power and sex among apes.* New York: Harper & Row.

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.

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.

Weissner, P. (1982). Risk, reciprocity and social influences on !Kung San economics. In E. Leacock & R. B. Lee (Eds.), *Politics and history in band societies.* Cambridge: Cambridge University Press.

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

Wilkinson, G. S. (1990, February). Food sharing in vampire bats. *Scientific American,* 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., & Williams, D. C. (1957). Natural selection of individually harmful social adaptations among sibs with special reference to social insects. *Evolution, 11,* 32–39.

Wilson, E. O. (1971). *The insect societies.* Cambridge, MA: Harvard University Press.