In Michael S. Gazzaniga (Ed.),
The New Cognitive Neurosciences, III.
(pp. 1295-1308).
Cambridge, MA: MIT Press 2005

93 Social Exchange: The Evolutionary Design of a Neurocognitive System

LEDA COSMIDES AND JOHN TOOBY

ABSTRACT How functionally specialized is the evolutionary design of our neural circuitry? Neuropsychological and cognitive research on human reasoning about social exchange indicates that at least some neurocomputational adaptations are quite narrow in scope. Evolutionary game theory shows that social exchange—cooperation for mutual benefit—can evolve and persist only if the cognitive programs that cause it conform to a narrow and complex set of design specifications. These design features were tested for and found, in experiments that simultaneously falsified theories claiming that more domain-general cognitive procedures cause reasoning about social exchange. The complex pattern of functional and neural dissociations found reveal so close a fit between adaptive problem and computational solution that a neurocognitive specialization for reasoning about social exchange is implicated, including a subroutine for cheater detection. This subroutine develops precocially (by ages 3–4) and appears cross-culturally: hunter-horticulturalists in the Amazon detect cheaters as reliably as adults who live in advanced market economies. The computational specialization found in adults appears to have been built by developmental mechanisms that evolved for that function; its design, ontogenetic timetable, and cross-cultural distribution are not consistent with any known domain-general learning process. In sum, the system that causes reasoning about social exchange shows evidence of being a cognitive instinct (Pinker, 1994): it is complexly organized for solving a well-defined adaptive problem our ancestors faced in the past, it reliably develops in all normal human beings, it develops without any conscious effort and in the absence of explicit instruction, it is applied without any conscious awareness of its underlying logic, and it is functionally and neurally distinct from more general abilities to process information or behave intelligently.

Everything should be made as simple as possible, but no simpler.
—Albert Einstein

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

Designed for social exchange?

By exchanging benefits—goods, services, acts of help and kindness—people can make themselves better off than they were before. This very basic fact of human social life is easy to take for granted. But when placed in zoological perspective, social exchange stands out as a strange phenomenon whose existence requires explanation.

ZOOLOGICAL DISTRIBUTION Despite widespread investigation, social exchange (reciprocity, reciprocal altruism) has been reported in only a tiny handful of species, such as chimpanzees, baboons, lions, and vampire bats (see Dugatkin, 1997, and Hauser, in press, for contrasting views on the animal findings). Most species do not engage in this very useful form of mutual help.

In contrast, social exchange is a characteristic of our species as language or tool use. Not only is social exchange found in every documented culture, but it is a feature of virtually every human life within each culture, taking on a multiplicity of forms, such as returning favors, sharing food, reciprocal gift giving, market exchange, and extending acts of help with the (implicit) expectation that they will be reciprocated (Cashdan, 1989; Fiske, 1991; Gurven, 2002). Paleoanthropological evidence suggests that certain forms of social exchange were present in hominids at least two million years ago (Isaac, 1978), and its presence in other primates suggests it may be even more ancient than that.

The fact that social exchange is an ancient and pervasive feature of human social life, yet rare in other species, is informative. It means that the neurocognitive machinery necessary for social exchange exists in humans, but not in most animals. But what, exactly, is the nature of the neurocognitive machinery that enables exchange, and how specialized is it for this function?

Is social exchange a by-product of neural circuitry that causes one to reason logically? To think intelligently? To reason about all conditional rules? To reason about deontic rules—moral rules involving obligation and entitlement? Or does the ability to engage in social exchange require evolved mechanisms that were tailored by natural selection specifically for social exchange?

The research discussed in this chapter explores a simple hypothesis: that the evolved, species-typical design of the human mind includes computational adaptations specialized for reasoning about social exchange.

EVOLUTIONARY FUNCTION and DESIGN EVIDENCE Social exchange clearly produces beneficial effects for those who successfully engage in it. I offer to provide a benefit to you, contingent on your satisfying a requirement that I specify. I impose that requirement in the hope that your satisfying it will create a situation that benefits me in some way. These conditional agreements—social contracts—are offered and accepted in the expectation that they will be rewarding for each party.

This means that the neurocognitive system that enables social exchange is beneficial. But this is not sufficient for showing that it was designed by natural selection to produce social exchange. Social exchange may simply be a side effect of a system that was designed for some entirely different function. How can one tell?

To demonstrate that the neurocognitive system that enables social exchange is an adaptation for that function, design evidence is needed. Computational systems, whether in brains or in computers, are composed of design features: properties that exist because they solve computational problems well. Natural selection is a causal feedback process that retains and discards properties from a species' design on the basis of how well they solve adaptive problems (evolutionarily recurrent problems whose solution promotes reproduction). To show that a system is an adaptation that evolved for a particular function, one must demonstrate that its properties solve a well-specified adaptive problem in a well-engineered way (Williams, 1966; Dawkins, 1986; Tooby and Cosmides, 1992).

The expectation of a fit between problem and solution can also be used to discover facts previously unknown. From a good specification of a computational problem one can predict and then look for representations and procedures that solve that problem well. In the reseach described here, we applied that approach to social exchange by (1) examining the selection pressures that arise in social exchange, (2) developing a task analysis of the computational problems that must be solved by a brain that was sculpted by these selection pressures, (3) using neuropsychological and cognitive methods to test for the presence of computational units that appear well designed for solving these problems, and (4) empirically testing to see whether performance is better explained as the by-product of mechanisms designed to solve some different, larger, or more general class of problems.

Selection pressures and design features

Selection pressures favoring social exchange exist whenever one organism (the provisioner) can change the behavior of a target organism to the provisioner's advantage by making the target's receipt of a provisioned benefit *conditional* on the target acting in a required manner. This mutual provisioning of benefits, each conditional on the other's compliance, is what is meant by social exchange or reciprocation (Cosmides, 1985; Cosmides and Tooby, 1989; Tooby and Cosmides, 1996). In social exchange, individuals agree, either explicitly or implicitly, to abide by a particular *social contract*. For ease of explication, let us define a social contract as a conditional rule that fits the following template: "If you accept a benefit from *X*, then you must satisfy *X*'s requirement" (where *X* is an individual or set of individuals).

An evolutionarily stable strategy, or ESS, is a strategy (a decision rule) that can arise and persist in a population because it produces fitness outcomes greater than or equal to alternative strategies (Maynard Smith, 1982). Whatever the decision rules are that guide social exchange in humans, it is likely that they embody an ESS (because they would not exist unless they had outcompeted alternatives). By using game theory and conducting computer simulations of the evolutionary process, one can determine which strategies for engaging in social exchange are ESSs.

In such simulations, social exchange is usually modeled as a repeated Prisoners' Dilemma (Trivers, 1971; Axelrod and Hamilton, 1981; Boyd, 1988; but see Tooby and Cosmides, 1996). The results show that the behavior of cooperators must be generated by programs that perform certain specific tasks very well if they are to be evolutionarily stable (Cosmides, 1985; Cosmides and Tooby, 1989). Here, we will focus on one of these requirements—cheater detection. A *cheater* is an individual who fails to reciprocate: who accepts the benefit specified by a social contract without satisfying the requirement that provision of that benefit was made contingent upon.

The ability to reliably and systematically detect cheaters is a necessary condition for cooperation in the repeated Prisoners' Dilemma to be an ESS (e.g., Williams, 1966; Trivers, 1971; Axelrod and Hamilton, 1981; Axelrod, 1984; Boyd, 1988). To see this, consider the fate of a program that, because it cannot detect cheaters, bestows benefits on others unconditionally. These unconditional helpers will increase the fitness of any nonreciprocating design they meet in the population. But when a nonreciprocating design is helped, the unconditional helper never recoups the expense of helping: the helper design incurs a net fitness cost while conferring a net fitness advantage on a design that does not help. As a result, a population of unconditional helpers is easily invaded and eventually outcompeted by designs that accept the benefits helpers bestow without reciprocating them.

In contrast, program designs that cause *conditional* helping—that help those who reciprocate the favor, but not those who fail to reciprocate—can invade a population of nonreciprocators and outcompete them. Moreover, a population of such designs can resist invasion by designs that do not nonreciprocate (cheater designs). Therefore, conditional helping, which requires the ability to detect cheaters, is an ESS.

Based on ESS analyses and the behavioral ecology of hunter-gatherers, one can define some of the computational requirements of an evolutionarily stable program for engaging in social exchange (Cosmides, 1985; Cosmides and Tooby, 1989). This task analysis of the required computations is what we mean by social contract theory. Social contract theory provides a standard of good design for this domain. That is, well-designed programs for engaging in social exchange should include features that execute the computational requirements specified in social contract theory.

Among the design features predicted by social contract theory are D1–D6:

D1. Social exchange is cooperation for mutual *benefit.* If there is nothing in a conditional rule that can be interpreted as a rationed benefit, then interperative procedures should not classify that rule as a social contract.

D2. Cheating is a specific way of violating a social contract: it is taking the benefit when you are not entitled to do so. Consequently, the cognitive architecture must define the concept of *cheating* using contentful representational primitives, referring to illicitly taken *benefits.* This implies that a system designed for cheater detection will not know what to look for if the rule specifies no benefit to the potential violator.

D3. The definition of cheating is also perspective dependent, because the item or action that one party views as a benefit is viewed as a requirement by the other party. The system needs to be able to compute a cost-benefit representation from the perspective of each participant, and define cheating with respect to that perspective-relative representation.

D4. To be an ESS, a design for conditional helping must not be outcompeted by alternative designs. Accidents and innocent mistakes that result in an individual being cheated are not markers of a design difference. A cheater detection system should look for cheaters: designs that cheat by intention rather than by accident. (Mistakes that result in one being cheated are relevant only insofar as they may not be true mistakes.)

D5. However unfamiliar the situation may be, rules that fit the template of a social contract should elicit cheater detection.

D6. Inferences made about social contracts should not follow the rules of a content-free formal logic. They should follow a content-specific adaptive logic, evolutionarily tailored for the domain of social exchange (described in Cosmides and Tooby, 1989).

Not only does cheating involve the violation of a conditional rule, but it involves a particular *kind* of violation of a particular *kind* of conditional rule. The rule must fit the template for a social contract; the violation must be one in which an individual intentionally took what that individual considered to be a benefit, and did so without satisfying the requirement.

Formal logics (e.g., the propositional calculus) are content-free; the definition of violation in standard logics applies to all conditional rules, whether they are social contracts, threats, or descriptions of how the world works. But, as we will see later, the definition of cheating implied by design features D1–D4 does not map onto this content-free definition of violation. What counts as cheating in social exchange is so content dependent that a detection mechanism equipped only with a domain-general definition of violation would not be able to solve the problem of cheater detection. This suggests that there should be a program specialized for cheater detection. To operate, this would have to function as a subcomponent of a system that, because of its domain-specialized structure, is well designed for detecting social conditionals involving exchange, interpreting their meaning, and successfully solving the inferential problems they pose: social contract algorithms.

Conditional reasoning and social exchange

Reciprocation is, by definition, social behavior that is conditional: you agree to deliver a benefit conditionally (conditional on the other person doing what you required in return). Understanding it therefore requires conditional reasoning.

Because engaging in social exchange requires conditional reasoning, investigations of conditional reasoning can be used to test for the presence of social contract algorithms. The hypothesis that the brain contains social contract algorithms predicts a dissociation in reasoning performance by content: a sharply enhanced ability to reason adaptively about conditional rules when those rules specify a social exchange. The null hypothesis is that there is nothing specialized in the brain for social exchange. This null hypothesis follows from the traditional assumption that reasoning is caused by content-independent processes. It predicts no enhanced conditional reasoning performance specifically triggered by social exchanges as compared to other contents.

A standard tool for investigating conditional reasoning is the Wason selection task, which asks one to look for potential violations of a conditional rule of the form *If P then Q*

Ebbinghaus disease was recently identified and is not yet well understood. So an international committee of physicians who have experience with this disease were assembled. Their goal was to characterize the symptoms, and develop surefire ways of diagnosing it.

Patients afflicted with Ebbinghaus disease have many different symptoms: nose bleeds, headaches, ringing in the ears, and others. Diagnosing it is difficult because a patient may have the disease, yet not manifest all of the symptoms. Dr. Buchner, an expert on the disease, said that the following rule holds:

"If a person has Ebbinghaus disease, then that person will be forgetful."
If P then Q

Dr. Buchner may be wrong, however. You are interested in seeing whether there are any patients whose symptoms violate this rule.

The cards below represent four patients in your hospital. Each card represents one patient. One side of the card tells whether or not the patient has Ebbinghaus disease, and the other side tells whether or not that patient is forgetful.

Which of the following card(s) would you definitely need to turn over to see if any of these cases violate Dr. Buchner's rule: "If a person has Ebbinghaus disease, then that person will be forgetful." Don't turn over any more cards than are absolutely necessary.

has Ebbinghaus disease	does not have Ebbinghaus disease	is forgetful	is not forgetful
P	*not-P*	*Q*	*not-Q*

FIGURE 93.1 The Wason selection task (descriptive rule, familiar content). In a Wason task, there is always a rule of the form *If P then Q*, and four cards showing the values *P*, *not-P*, *Q*, and *not-Q* (respectively) on the side that the subject can see. From a logical point of view, only the combination of *P* and *not-Q* can violate this rule, so the correct answer is to check the *P* card (to see if it has a *not-Q* on the back), the *not-Q* card (to see if it has a *P* on the back), and no others. Few subjects answer correctly, however, when the conditional rule is descriptive (indicative), even when its content is familiar. For example, only 26% of subjects answered the above problem correctly (by choosing "has Ebbinghaus disease" and "is not forgetful"). Most choose either *P* alone, or *P&Q*. (The italicized *P*s and *Q*s are not in problems given to subjects.)

(Wason, 1966, 1983; Wason and Johnson-Laird, 1972). Using this task, an extensive series of experiments has been conducted that address the following questions:

1. Do our minds include cognitive machinery that is specialized for reasoning about social exchange? (alongside some other domain-specific mechanisms, each specialized for reasoning about a different adaptive domain involving conditional behavior?) Or,

2. Is the cognitive machinery that causes good conditional reasoning general—does it operate well regardless of content?

If the human brain had cognitive machinery that causes good conditional reasoning regardless of content, then people should be good at tasks requiring conditional reasoning. For example, they should be good at detecting violations of conditional rules. Yet studies with the Wason selection task show that they are not. The Wason task in figure 93.1 is illustrative. The correct answer (choose *P*, choose *not-Q*) would be intuitively obvious if our minds were equipped with reasoning procedures specialized for detecting *logical* violations of conditional rules. But this is not obvious. Studies in many nations have shown that reasoning performance on descriptive (indicative) rules like this is low: only 5%–30% of people give the logically correct answer, even when the rule involves familiar terms drawn from everyday life (Wason, 1966, 1983; Manktelow and Evans, 1979; Cosmides, 1989; Sugiyama, Tooby, and Cosmides, 2002).

A DISSOCIATION BY CONTENT People are poor at detecting violations of conditional rules when their content is descriptive. Does this result generalize to conditional rules that express social contracts? No. People who ordinarily cannot detect violations of if-then rules can do so easily and accurately when that violation represents cheating in a situation of social exchange. This pattern—good violation detection for social contracts but not for descriptive rules—is a dissociation in reasoning elicited by differences in the conditional rule's content. It provides (initial) evidence that the mind has reasoning procedures specialized for detecting cheaters.

More specifically, when asked to look for violations of a conditional rule that fits the social contract template—"If you take benefit B, then you must satisfy requirement R" (e.g., "If you borrow my car, then you have to fill up the tank

A.

Teenagers who don't have their own cars usually end up borrowing their parents' cars. In return for the privilege of borrowing the car, the Carter's have given their kids the rule,

"If you borrow my car, then you have to fill up the tank with gas."

Of course, teenagers are sometimes irresponsible. You are interested in seeing whether any of the Carter teenagers broke this rule.

The cards below represent four of the Carter teenagers. Each card represents one teenager. One side of the card tells whether or not a teenager has borrowed the parents' car on a particular day, and the other side tells whether or not that teenager filled up the tank with gas on that day.

Which of the following card(s) would you definitely need to turn over to see if any of these teenagers are breaking their parents' rule: "If you borrow my car, then you have to fill up the tank with gas." Don't turn over any more cards than are absolutely necessary.

borrowed car	did not borrow car	filled up tank with gas	did not fill up tank with gas

B.

The mind translates social contracts into representations of benefits and requirements, and it inserts concepts such as "entitled to" and "obligated to", whether they are specified or not.

How the mind "sees" the social contract above is shown in ***bold italics***.

"If you borrow my car, then you have to fill up the tank with gas."

If you take the benefit, then you are obligated to satisfy the requirement.

borrowed car	did not borrow car	filled up tank with gas	did not fill up tank with gas
= accepted the benefit	***= did not accept the benefit***	***= satisfied the requirement***	***= did not satisfy the requirement***

FIGURE 93.2 Wason task with a social contract rule. (*A*) In response to this social contract problem, 76% of subjects chose *P* & *not-Q* ("borrowed the car" and "did not fill the tank with gas"), the cards that represent potential cheaters. Yet only 26% chose this (logically correct) answer in response to the descriptive rule in figure 93.1. Although this social contract rule involves familiar items, unfamiliar social contracts elicit the same high performance. (*B*) How the mind represents the social contract shown in *A*. According to inferential rules specialized for social exchange (but not according to formal logic), "If you take the benefit, then you are obligated to satisfy the requirement" implies "If you satisfy the requirement, then you are entitled to take the benefit." Consequently, the rule in *A* implies: "If you fill the tank with gas, then you may borrow the car" (see figure 93.4, switched social contracts).

with gas")—people check the individual who accepted the benefit (borrowed the car; *P*) and the individual who did not satisfy the requirement (did not fill the tank; *not-Q*), that is, the cases that represent potential cheaters (figure 93.2*A*). The adaptively correct answer is immediately obvious to most subjects, who commonly experience a pop-out effect. No formal training is needed. Whenever the content of a problem asks one to look for cheaters in a social exchange, subjects experience the problem as simple to solve, and their performance jumps dramatically. In general, 65%–80% of subjects get it right, the highest performance found for a task of this kind (for reviews, see Cosmides, 1985, 1989; Cosmides and Tooby, 1992, 1997; Gigerenzer and Hug, 1992; Platt and Griggs, 1993; Fiddick, Cosmides, and Tooby, 2000).

Given the content-free syntax of formal logic, investigating the person who borrowed the car (*P*) and the person who did not fill the gas tank (*not-Q*) is logically equivalent to investigating the person in figure 93.1 with Ebbinghaus disease (*P*) and the person who is not forgetful (*not-Q*). But everywhere it has been tested (adults in the United States, United Kingdom, Germany, Italy, France, Hong Kong, Japan; schoolchildren in Quito, Ecuador; Shiwiar hunter-horticulturalists in the Ecuadorian Amazon), people do not treat social exchange problems as equivalent to other kinds of reasoning problems (Cheng and Holyoak, 1985; Cosmides, 1989; Platt and Griggs, 1993; Hasegawa and Hiraishi, 2000; Sugiyama, Tooby, and Cosmides, 2002; supports D5, D6). Their minds distinguish social exchange contents, and reason as if they were translating these situations into representational primitives such as ***benefit***, ***cost***, ***obligation***, ***entitlement***, ***intentional***, and ***agent*** (figure 93.2*B*; Cosmides and Tooby, 1992; Fiddick, Cosmides, and Tooby, 2000). Reasoning problems and their elements could be sorted into indefinitely many categories based on their content or structure (including the propositional calculus's two content-free categories, antecedent and consequent). Yet even in remarkably different cultures, the same mental categorization occurs. This cross-culturally recurrent dissociation by content was predicted by social contract theory's adaptationist analysis.

In the next section we review experiments conducted to test for design features that should be present in a system specialized for social exchange. Each experiment testing for a design feature was also constructed to pit the adaptive specialization hypothesis against at least one alternative by-product hypothesis, so design feature and by-product implications will be discussed in tandem. As we will show, reasoning performance on social contracts is not explained by familiarity effects, by a content-free formal logic, by a permission schema, or by a general deontic logic (table 93.1).

Do unfamiliar social contracts elicit cheater detection (D5)?

One needs to understand each new opportunity to exchange as it arises, so it was predicted that social exchange reasoning should operate even for unfamiliar social contract rules (D5). (This distinguishes social contract theory strongly from theories that explain reasoning performance as the product of general learning strategies plus experience: The most natural prediction for such skill acquisition theories is that performance should be a function of familiarity.) Surprisingly, social contract theory is supported: cheater detection occurs even when the social contract is wildly unfamiliar (figure 93.3*A*). For example, the rule "If a man eats cassava root, then he must have a tattoo on his face" can be made to fit the social contract template by explaining that the people involved consider eating cassava root to be a benefit (the rule then implies that having a tattoo is the requirement one must satisfy to be eligible for that benefit). When this is done, this outlandish, culturally alien rule elicits the same high level of cheater detection as highly familiar social exchange rules (Cosmides, 1985, 1989; Gigerenzer and Hug, 1992; Platt and Griggs, 1993).

TABLE 93.1

Alternative (by-product) hypotheses eliminated

B1	That familiarity can explain the social contract effect
B2	That social contract content merely activates the rules of inference of the propositional calculus (logic)
B3	That any problem involving payoffs will elicit the detection of logical violations
B4	That permission schema theory can explain the social contract effect
B5	That social contract content merely promotes "clear thinking"
B6	That a content-independent deontic logic can explain social contract reasoning
B7	That a single mechanism operates on all deontic rules involving subjective utilities
B8	That relevance theory can explain social contract effects (see also Fiddick et al., 2000)

ELIMINATING FAMILIARITY (B1) The dissociation by content—good performance for social contract rules but not for descriptive ones—has nothing to do with the familiarity of the rules tested. Surprisingly, familiarity is neither necessary nor sufficient for eliciting high performance (B1 of table 93.1).

First, familiarity does not produce high levels of performance for descriptive rules (Manktelow and Evans, 1979; Cosmides, 1989). For example, the Ebbinghaus problem in figure 93.1 involves a familiar causal relationship, a disease causing a symptom, embedded in a real-world context. Yet only 26% of 111 college students that we tested produced the logically correct answer, *P & not-Q*, for this problem. If familiarity fails to elicit high performance on descriptive rules, then it also fails as an explanation for high performance on social contracts.

Second, the fact that unfamiliar social contracts elicit high performance shows that familiarity is not necessary for eliciting violation detection. Third, and most surprising, people are just as good at detecting cheaters on culturally unfamiliar or imaginary social contracts as they are at detecting cheaters on completely familiar social contracts (Cosmides, 1985). This provides a challenge for any counterhypothesis resting on a general-learning skill acquisition account (most of which rely on familiarity and repetition).

Adaptive logic, not formal logic (D3, D6)

Social contract problems can be constructed so as to elicit a logically correct answer (*P & not-Q*; see figure 93.2*A*). This has led some to conclude that social exchange content simply activates a dormant content-free logical faculty. But this is not the case.

Good cheater detection is not the same as good detection of logical violations (and vice versa). Hence, problems can be created in which the search for cheaters will result in a logically incorrect response (and the search for logical violations will fail to detect cheaters; figure 93.4). When given such problems, people look for cheaters, thereby giving a logically incorrect answer (*Q & not-P*). Experiments with perspective change and switched social contracts provide examples.

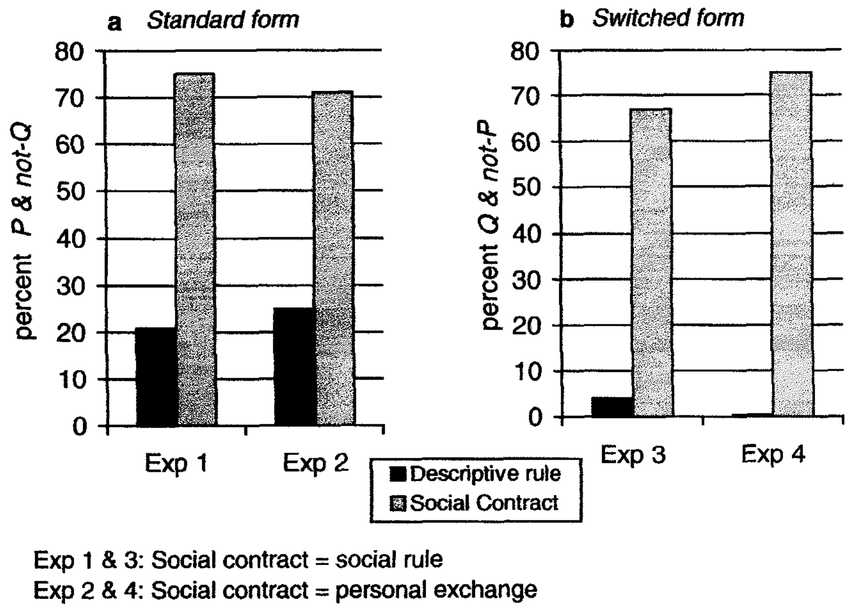

FIGURE 93.3 Detecting violations of unfamiliar conditional rules: social contracts versus descriptive rules. In these experiments, the same unfamiliar rule was embedded either in a story that caused it to be interpreted as a social contract or in a story that caused it to be interpreted as a rule describing some state of the world. For social contracts, the correct answer is always to pick the *benefit accepted* card and the *requirement not satisfied* card. (*A*) For standard social contracts, these correspond to the logical categories *P* and *not-Q*. *P & not-Q* also happens to be the logically correct answer. More than 70% of subjects chose these cards for the social contracts, but fewer than 25% chose them for the matching descriptive rules. (*B*) For switched social contracts, the *benefit accepted* and *requirement not satisfied* cards correspond to the logical categories *Q* and *not-P*. This is not a logically correct response. Nevertheless, about 70% of subjects chose it for the social contracts; virtually no one chose it for the matching descriptive rule (see figure 93.4).

PERSPECTIVE CHANGE As predicted (D3), the mind's automatically deployed definition of cheating is tied to the perspective one is taking (Gigerenzer and Hug, 1992). For example, consider the following social contract:

(1) *If an employee is to get a pension, then that employee must have worked for the firm for over 10 years.*

This rule elicits different answers depending on whether subjects are cued into the role of employer or employee. Those in the employer role look for cheating by employees, investigating cases of *P* and *not-Q* (employees with pensions; employees who have worked for fewer than 10 years). Those in the employee role look for cheating by employers, investigating cases of *not-P* and *Q* (employees with no pension; employees who have worked more than 10 years). *Not-P & Q* is correct if the goal is to find out whether the employer is a cheater. But it is not *logically* correct. Content-free logical rules would always look for the co-occurrence of *P* and *not-Q*; perspective, a content-rich concept, is irrelevant to logic.

SWITCHED SOCIAL CONTRACTS Assume you are the employer looking for cheating by employees. You are looking for violations of this social contract:

(2) *If an employee has worked for the firm for over 10 years, then that employee gets a pension.*

The mind recognizes (1) and (2) as expressing the same social contract (figures 93.2*B*, 93.4). For (2), as for (1), finding employees who cheat involves checking the employee who took the benefit (the pension) without meeting the requirement (worked < 10 years). But now these fall into the logical categories *not-P* and *Q*. When given social contracts with the benefit in the consequent clause (a "switched" format), subjects overwhelmingly choose *Q* & *not-P*—an answer that is adaptively correct but logically incorrect (Cosmides, 1989; Gigerenzer and Hug, 1992; supports D2, D6) (fig. 93.3*B*).

ELIMINATING LOGIC (B2, B3) In these experiments, people did not follow the inferential rules of a content-free logic; by doing so they would have failed to detect cheaters (see figure 93.4). They applied inferential rules specific to social exchange, and therefore detected cheaters. The results show that performance on social contract problems is not caused by the activation of a dormant logic faculty (also see Fiddick et al., 2000).

In fact, social contract reasoning can be maintained in the face of impairments in general logical reasoning. Individuals with schizophrenia manifest deficits on virtually any test of general intellectual functioning they are given (McKenna, Clare, and Baddeley, 1995), yet their ability to detect cheaters can remain intact. Maljkovic (1987) tested the rea-

Consider the following rule:

Standard format:

If you take the ***benefit****, then satisfy my* ***requirement*** (e.g., "If I give you $50, then give me your watch.")

If P then Q

Switched format:

If you satisfy my ***requirement****, then take the* ***benefit*** (e.g., "If you give me your watch, then I'll give you $50.")

If P then Q

The cards below have information about four people. Each card represents one person. One side of a card tells whether the person accepted the benefit, and the other side of the card tells whether that person satisfied the requirement. Indicate only those card(s) you definitely need to turn over to see if any of these people have violated the rule.

	✔ Benefit accepted	Benefit not Accepted	Requirement satisfied	✔ Requirement not satisfied
Standard:	*P*	*not-P*	*Q*	*not-Q*
Switched:	*Q*	*not-Q*	*P*	*not-P*

FIGURE 93.4 Generic structure of a Wason task when the conditional rule is a social contract. A social contract can be translated into either social contract terms (benefits and requirements) or logical terms (*P*s and *Q*s). Checkmarks indicate the correct card choices if one is looking for cheaters; these cards should be chosen by a cheater detection subroutine, whether the exchange was expressed in a standard or switched format. This results in a logically incorrect answer (*Q & not-P*) when the rule is expressed in the switched format and a logically correct answer (*P & not-Q*) when the rule is expressed in the standard format. By testing switched social contracts, one can see that the reasoning procedures activated cause one to detect cheaters, not logical violations (see figure 93.3*B*). A logically correct response to a switched social contract, where *P* = *requirement satisfied* and *not-Q* = *benefit not accepted*, would fail to detect cheaters.

soning of patients exhibiting positive symptoms of schizophrenia and compared their performance with that of hospitalized control patients. Compared to the control patients, the schizophrenic patients were impaired on more general (non-Wason) tests of logical reasoning, in a way typical of individuals with frontal lobe dysfunction. But their ability to detect cheaters on Wason tasks was unimpaired: it was indistinguishable from that of the controls, and showed the typical dissociation by content. This selective preservation of social exchange reasoning is consistent with the notion that reasoning about social exchange is handled by a dedicated system that can operate even when the systems responsible for more general reasoning are damaged.

Benefits are necessary for cheater detection (D1, D2)

The function of a social exchange for each participant is to gain access to a benefit that would otherwise be unavailable to them. Therefore, an important cue that a conditional rule is a social contract is the presence in it of a desired benefit under the control of an agent.

In social exchange, this agent *permits* you to take a benefit from him or her, conditonal upon your having met the agent's requirement. There are, however, many situations other than social exchange in which an action is permitted conditionally. A *permission rule* is any deontic conditional that fits the template "If one is to take action A, then one must satisify precondition R"[1] (table 93.2; Cheng and Holyoak, 1985, 1989). All social contracts are permission rules, but there are many permission rules that are not social contracts (see table 93.2, figure 93.5). *Permission schema theory* proposes that the reasoning system that causes cheater detection is not specialized for that purpose. According to this theory, good violation detection is elicited by the entire class of permission rules—a far more inclusive and general set (Cheng and Holyoak, 1985, 1989).

TABLE 93.2

*The four production rules of the permission schema**

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.

Social contracts and precautions fit the template of Rule 1:

If the benefit is to be taken, then the requirement must be satisfied.

If the hazardous action is to be taken, then the precaution must be taken.

* Permission schema of Cheng and Holyoak (1985).

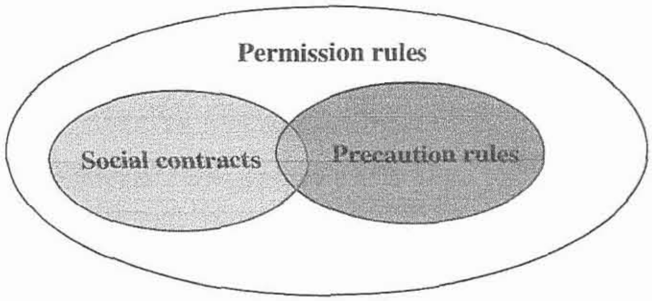

FIGURE 93.5 The class of permission rules is larger than, and includes, social contracts and precautionary rules. Many of the permission rules we encounter in everyday life are neither social contracts nor precautions (white area). Rules of civil society (etiquette, customs, traditions), bureaucratic rules, corporate rules—many of these are conditional rules that do not regulate access to a benefit or involve a danger. Permission schema theory (see table 93.2) predicts high performance for all permission rules; however, permission rules that fall into the white area do not elicit the high levels of performance that social contracts and precaution rules do. Neuropsychological and cognitive tests show that performance on social contracts dissociates from other permission rules (white area), from precautionary rules, and from the general class of deontic rules involving subjective utilities. These dissociations would be impossible if reasoning about social contracts and precautions were caused by a single schema that is general to the domain of permission rules.

Just how precise and functionally specialized is the reasoning system that causes cheater detection? Permission schema theory predicts uniformly high performance for all permission rules, whether they are social contracts or not. In contrast, social contract theory predicts dissociations *within* the class of permission rules: being a permission rule will not be sufficient, and large subsets of permission rules will fail to elicit the effect. For example, according to social contract theory, removing benefits (D1, D2) and/or intentionality (D4) from a social contract will result in a permission rule that does not elicit violation detection.

The benefit prediction was tested by Cosmides and Tooby (1992), who constructed Wason tasks involving a fictitious culture in which the elders made laws governing the conditions under which adolescents were permitted to take certain actions. For all tasks, the law fit the template for a permission rule. What varied was whether the action to be taken was a benefit or an unpleasant chore.

A cheater detection subroutine looks for benefits illicitly taken; without a benefit, it doesn't know what kind of violation to look for (D1, D2). When the permitted action was a benefit, 80% of subjects answered correctly; when it was a chore, only 44% did so. This dramatic decrease in violation detection was predicted in advance by social contract theory; in contrast, it violates the central prediction of permission schema theory, that being a permission rule is sufficient to facilitate violation detection. For similar results, see Manktelow and Over (1991), Platt and Griggs (1993), and Barrett (1999).

This dissociation within the domain of permission rules supports the psychological reality of social contract categories; it shows that the representations necessary to trigger differential reasoning are more content-specific than those of the permission schema.

Social contract violations must be intentional (D4)

Evolutionarily, the function of a cheater detection subroutine is to correctly connect an attributed disposition (to cheat) with a person (a cheater), not simply to recognize instances wherein an individual did not get what she was entitled to. This is because the fitness benefit of cheater detection is the ability to avoid squandering costly future cooperative efforts on those who will not reciprocate. Violations of social contracts are relevant only insofar as they reveal individuals disposed to cheat—individuals who cheat by design, not by accident. Noncompliance caused by factors other than disposition, such as accidental violations and other innocent mistakes, do not reveal the disposition or design of the exchange partner; they may result in someone being cheated, but without indicating the presence of a cheater. Therefore, social contract theory predicts another additional level of cognitive specialization beyond detecting compliance or noncompliance with a social contract. The subroutine should look for *intentional* violations (D4).

Given the same social contract rule, one can manipulate contextual factors to change the nature of the violation from intentional cheating to an innocent mistake. One experiment, for example, compared a condition in which the potential rule violator was well-meaning but inattentive to one in which she had an incentive to intentionally cheat. Varying intentionality caused a radical change in performance, from 68% correct in the intentional cheating condition to 27% correct in the innocent mistake condition (Cosmides, Barrett, and Tooby, 2004; supports D4; disconfirms B1–B8). Fiddick (1998, 2004) found the same effect (as did Gigerenzer and Hug, 1992, using a different context manipulation).

Barrett (1999) conducted a series of parametric studies to find out whether the drop in performance in the innocent mistake condition was caused by the violator's lack of intentionality or by her inability to benefit from her mistake (see D2). He found that both factors contributed, independently and additively, to the drop.

ELIMINATING PERMISSION SCHEMA THEORY (B4) Cheng and Holyoak (1985, 1989) speculate that repeated encounters with permission rules cause domain-general learning mechanisms to induce a *permission schema*, consisting of four production rules (see table 93.2). This schema generates inferences about any conditional rule that fits the permission rule template, resulting in good violation detection for permission rules. However, permission schema theory cannot explain the above results.

According to permission schema theory, (1) all permission rules should elicit high levels of violation detection, whether the action term is a benefit or a chore; and (2) all permission rules should elicit high levels of violation detection, whether the violation was committed intentionally or accidentally. Both predictions fail. Permission rules fail to elicit high levels of violation detection when the action term is neutral or unpleasant (yet not hazardous; see later discussion). Moreover, people are poor at detecting accidental violations of social contract rules (which are a species of permission rule; see figure 93.5). Taken together, these results cast doubt on the hypothesis that the mind contains or develops a permission schema of the kind postulated by Cheng and Holyoak (1985, 1989).

The same results also falsify B6, that cheater detection on social contracts is caused by a content-free deontic logic (for discussion, see Manktelow and Over, 1987), as well as a proposal by Fodor (2000). All the rules tested above were deontic rules, but not all elicited violation detection. B5—that social contract rules elicit good performance because we understand their implications (e.g., Almor and Sloman, 1996)—is eliminated by the intention versus accident dissociation (the same social contract failed to elicit violation detection in the accident condition).

In short, it is not enough to admit that moral reasoning or social reasoning is special; the specificity implicated is far narrower in scope.

A neuropsychological dissociation between social contracts and precautions

The notion of multiple adaptive specializations is commonplace in physiology: the body is composed of many organs, each designed for a different function. Nevertheless, many cognitive neuroscientists are skeptical of multiple evolved specializations when these involve high-level cognitive operations such as reasoning. From an evolutionary perspective, however, social contracts are not the only conditional rules for which we should have specialized mechanisms (e.g., Tooby and Cosmides, 1989, on threats). Alongside specializations for reasoning about social exchange and threats, the human cognitive architecture should contain computational machinery specialized for managing hazards that causes good violation detection on precautionary rules. A system well designed for reasoning about hazards and precautions should have properties different from one for detecting cheaters (some of which have been tested for and found; Fiddick, 1998, 2004; Fiddick, Cosmides, and Tooby, 2000; Stone et al., 2002).

Precautionary rules are conditional rules that fit the template, *If one is to engage in hazardous activity H, then one must take precaution R* (e.g., "If you are working with toxic gases, then you must wear a gas mask"; Fiddick, Cosmides, and Tooby, 2000; Stone et al., 2002). Being able to detect when someone is in danger from having violated a precautionary rule is of obvious adaptive value. Tests with the Wason task show that precautionary rules strongly elicit the search for potential violators: subjects look for people who have engaged in the hazardous activity without taking the appropriate precaution (e.g., the "worked with toxic gases" card (*P*) and the "did not wear gas mask" card (*not-Q*)).

Like social contracts, precautionary rules are conditional, deontic, involve subjective utilities, and have the same pragmatic implications for Wason tasks (see table 93.2). Moreover, people are as good at detecting violators of precautionary rules as they are at detecting cheaters on social contracts. This has led some to conclude that reasoning about social contracts and precautions is caused by a single more general mechanism (e.g., general to permissions or to deontic rules involving subjective utilities; Manktelow and Over, 1988, 1990, 1991; Cheng and Holyoak, 1989; Sperber, Cara, and Girotto, 1995).

One Mechanism or Two? If reasoning about social contracts and precautions is caused by a single mechanism, then neurological damage to this mechanism should lower performance on both rules. But if reasoning about these two domains is caused by two functionally distinct mechanisms, then it is possible for social contract algorithms to be damaged while leaving precautionary mechanisms unimpaired, and vice versa.

Stone and colleagues (2002) developed a battery of Wason tasks that tested social contracts, precautionary rules, and descriptive rules. The social contracts and precautionary rules elicited equally high levels of violation detection from normal subjects (who got 70% and 71% correct, respectively). For each subject, a difference score was calculated: percent correct for precautions minus percent correct for social contracts. For normal subjects, these difference scores were close to zero (mean = 1.2 percentage points, SD = 11.5).

Stone and colleagues (2002) administered this battery of Wason tasks to R.M., a patient with bilateral damage to his medial orbitofrontal cortex and anterior temporal cortex (which had disconnected both amygdalae). R.M.'s performance on the precaution problems was 70% correct: equivalent to that of the normal controls. In contrast, his performance on the social contract problems was only 39% correct. Whereas the average difference score for control subjects was 1.2, R.M.'s difference score (precaution minus social contract) was 31 percentage points. This is 2.7 SD larger than the control mean ($P < 0.005$).

Double dissociations are helpful in ruling out differences in task difficulty as a counterexplanation for a given dissociation (Shallice, 1988), but here the tasks were perfectly matched for difficulty. The social contracts and precautionary rules given to R.M. were logically identical, posed identical task demands, and were equally difficult for normal subjects. Moreover, because the performance of the normal controls was not at ceiling, ceiling effects could not be masking real differences in the difficulty of the two sets of problems. In this case, a single dissociation is telling.

R.M.'s dissociation supports the hypothesis that reasoning about social exchange is caused by a different computational system than reasoning about precautionary rules—a two-mechanism account.

Although tests of this kind cannot conclusively establish the anatomical location of a mechanism, tests with other patients suggest that amygdalar disconnection was important in creating this selective deficit.[2]

Eliminating One Mechanism Hypotheses (B6–B8; B1–B4) Every alternative explanation of cheater detection proposed so far claims that reasoning about social contracts and precautions is caused by the same computational system. R.M.'s dissociation is inconsistent with these one-mechanism accounts. These include mental logic (Rips, 1994), mental models (Johnson-Laird and Byrne, 1991), decision theory/optimal data selection (Kirby, 1994; Oaksford and Chater, 1994), permission schema theory (Cheng and Holyoak, 1989), relevance theory (Sperber, Cara, and Girotto, 1995[3]), and Manktelow and Over's (1991) view implicating a system that is general to any deontic rule that involves subjective utilities.

Indeed, no other reasoning theory even distinguishes between precautions and social contract rules; the distinction is derived from evolutionary-functional analyses, and is purely in terms of content. These results indicate the presence of a very narrow, content-based cognitive specialization within human reasoning.

The development of social contract reasoning

The evidence strongly supports the claim that reasoning about social exchange is caused by computational machinery that is specialized for this function in adults: in other words, social contract algorithms. But how was this computational specialization produced? Do humans have domain-specific mechanisms that are designed to cause the development of social contract algorithms? Or are they the outcome of a domain-general learning process?

Precocious Development of Cheater Detection Children understand what counts as cheating on a social contract by age 3 (Harris and Núñez, 1996; Núñez and Harris, 1998b; Harris, Núñez, and Brett, 2001). This has been shown repeatedly in experiments by Harris and Núñez using an evaluation task, in which the child must identify the picture in which a character is violating the rule. For social contracts, British 3-year-olds chose the correct picture 72%–83% of the time and 4-year-olds chose correctly 77%–100% of the time (Harris and Núñez, 1996; Núñez and Harris, 1998a; Harris, Núñez, and Brett, 2001). The same effects were found for preschoolers from the United Kingdom, Colombia, and (with minor qualifications) from rural Nepal.

The performance of the preschoolers was adult-like in other ways. Like adults, the preschoolers did well whether the social contract was familiar or unfamiliar, and understood that taking the benefit was conditional on meeting the requirement. Also like adults, intentionality mattered to the children: intentional violations were viewed as "naughty" far more often than accidental ones (80% vs. 10% by age 4; 65% vs. 17% at age 3; Núñez and Harris, 1998a). Moreover, the children tested by Harris and Núñez (1996) showed the same dissociation between social contract and descriptive rules as adults: 3–4-year-olds chose the correct violation condition only 40% of the time for descriptive rules but 72%–83% of the time for social contracts. By age 5, children could solve the full array, four-card Wason task when the conditional rule expressed a social contract (Núñez and Harris, 1998b).

Cross-Cultural Invariances and Dissociations The ESS concept carries predictions about development. Because detecting cheaters is necessary for social exchange to maintain itself in an evolving species—to be an ESS (D4)—its development should be buffered against cultural variation. The hypothesis that social exchange reasoning is caused by an evolved specialization therefore predicts that cheater detection will be found in all human cultures. In contrast, the development of ESS-irrelevant aspects of Wason performance are under no selection to be uniform across cultures.

Sugiyama, Tooby, and Cosmides (2002) tested this prediction among the Shiwiar, a hunter-horticultural population in a remote part of the Ecuadorian Amazon. Good cheater detection had already been established in the United States, Europe, Hong Kong, and Japan. But adults in advanced market economies engage in more trade, especially with strangers, than people who hunt and garden in remote parts of the Amazon. Anonymity facilitiates cheating; markets increase the volume of transactions experienced by each individual. If no evolved specialization is involved—that is, if the mechanism is induced through repeated experience with cheating—then it may not be found outside the industrialized world.

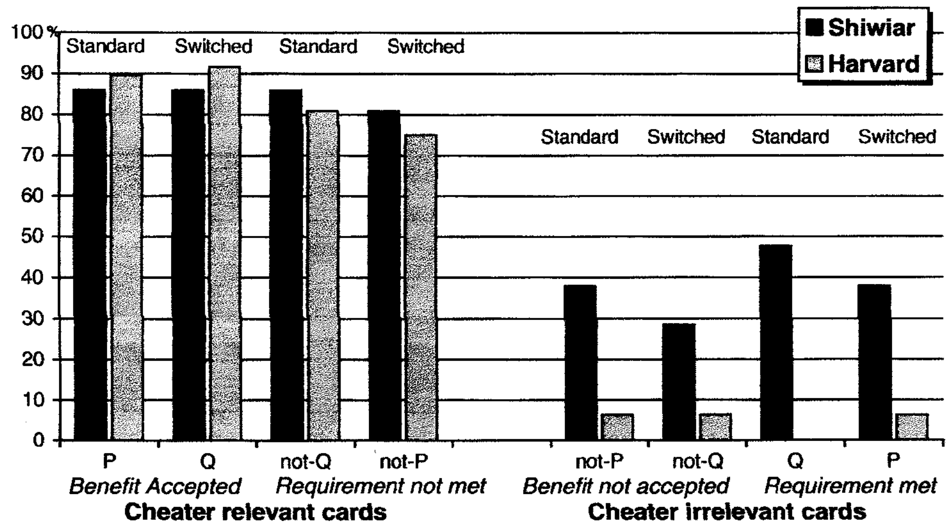

FIGURE 93.6 Performance of Shiwiar hunter-horticulturalists and Harvard undergraduates on standard and switched social contracts. Graphed is the percentage of subjects choosing each card. There was no difference between the two populations in their choice of cheater-relevant cards (*benefit accepted, requirement not satisfied*). They differed only in their choice of cheater-irrelevant cards (Shiwiar showing some interest in cards that could reveal acts of generosity or fair play). Shiwiar high performance on cheater-relevant cards is not caused by indiscriminate interest in all cards. Holding logical category constant, Shiwiar always chose a card more frequently when it was relevant to cheater detection than when it was not. This can be shown by comparing performance on standard versus switched social contracts. (For example, the *P* card is cheater relevant for a standard social contract but not for a switched one; see figure 93.4.)

The Shiwiar live in a culture as different from that of Western subjects as any remaining on the planet. Yet the Shiwiar were just as good at detecting cheaters on Wason tasks (figure 93.6). For cheater-relevant cards, the performance of Shiwiar hunter-horticulturalists was identical to that of Harvard undergraduates. Shiwiar differed only in that they were slightly more likely to show interest in cheater-irrelevant cards, the ones that could reveal acts of generosity.

The Shiwiar results suggest that the brain mechanism responsible for cheater detection reliably develops even in disparate cultural contexts—just what one would expect of a universal feature of human nature. There was no dissociation between cultures in the parts of the mechanism necessary to its performing its evolved function. The only "cultural dissociation" was in ESS-irrelevant aspects of performance.

Conclusion: Does domain-general learning build this computational specialization?

Reasoning about social exchange narrowly dissociates from other forms of reasoning, both cognitively and neurally. It displays design features specially tailored to fit the computational requirements necessary to produce an evolutionarily stable form of conditional helping (as opposed to the many kinds of helping that are culturally encouraged).

However, many psychologists believe that high-level cognitive competences like this emerge from general-purpose cognitive abilities trained by culturally specific activities. Such domain-general accounts rely on experience, familiarity, and repetition as explanatory variables. But the counterhypothesis that social exchange reasoning developed through some form of domain-general learning runs into a series of difficulties:

1. There is no evidence that familiarity, experience, or repetition improves conditional reasoning in any domain.
2. Neither experience with type of rule nor familiarity with specific rules accounts for which reasoning skills develop and which do not. For example, humans do not appear to develop the ability to reason well about large classes of rules with which they have far more experience than they do with social exchanges. We know this because social exchanges are a small subset of (for example) conditional rules, deontic rules, and permission rules; so, by class inclusion, humans necessarily have far more experience with these more general domains of rules (caption, figure 93.5). It was on this basis that Cheng and Holyoak (1985, 1989) argued that domain-general inductive processes *should* produce the more abstract and inclusive permission schema rather than social contract algorithms. Yet careful tests show that the permission schema does not exist, nor any of the other more inclusive competences that this view predicts.
3. Preschoolers have a limited base of experience, yet in reasoning about social exchange, they show virtually all the features of special design that adults do. It is difficult to understand why a domain-general learning process would cause the early and uniform acquisition of reasoning in

this domain yet fail to do so for others (Cosmides and Tooby, 1997).

4. Culture is often invoked as a schema-building factor. Yet, despite a massive difference in experience with trade and cheating, there was no difference between Shiwiar and American adults in cheater detection.

5. That people are good at detecting intentional cheating but not accidental mistakes is a prediction of the evolutionary task analysis of exchange. It is specifically the detection of intentional cheaters that makes contingent exchange evolutionarily stable against exploitation by cheaters (i.e., an ESS). In contrast, nonevolutionary theories of the origin of social exchange reasoning should predict good violation detection regardless of cause. A single inference procedure that scrutinizes cases in which the benefit was accepted and the requirement was not met indiscriminately reveals both accidental and intentional violations. Both represent damage to personal utility, both are useful to know, and both require detection if the participant is to get what she wants and is entitled to. From a pragmatic, utility-based perspective, it represents a strange addition to the competence to have the ordinarily deployed procedure inactive when there is evidence that the mistake would not have been intentional.

6. Similarly, it is not clear how or why domain-general learning would cause a cultural dissociation in the ESS irrelevant aspects of Wason-based social exchange reasoning, but not in the ESS-relevant aspects of cheater detection.

In contrast, the hypothesis that social contract algorithms were built by a developmental process designed for that function neatly accounts for all the developmental facts: that cheater detection develops invariantly across widely divergent cultures (whereas other aspects dissociate); that social exchange reasoning and cheater detection develop precocially; that they operate smoothly regardless of experience and familiarity; that they detect cheating and not other kinds of violations; that the developmental process results in a social contract specialization rather than one for more inclusive classes such as permissions.

The simplest, most parsimonious explanation that can account for all the results—developmental, neuropsychological, and cognitive—is that social contract algorithms are a reliably developing component of a universal human nature, designed by natural selection to produce an evolutionarily stable strategy for conditional helping.

NOTES

1. Cheng and Holyoak (1985) also propose an obligation schema, but for most rules tested (especially social contracts), this leads to the same predictions as the permission schema (see Cosmides, 1989).

2. Stone and colleagues tested two other patients with overlapping but different patterns of brain damage. R.B. had more extensive bilateral orbitofrontal damage than R.M., but his right temporal pole was largely spared (thus he did not have bilateral disconnection of the amygdalae): his scores were 85% correct for precautions and 83% correct for social contracts. B.G. had extensive bilateral temporal pole damage compromising (though not severing) input into both amygdalae, but his orbitofrontal cortex was completely spared: he scored 100% on both sets of problems.

3. For a full account of the problems relevance theory has explaining social contract reasoning, see Fiddick and colleagues (2000).

REFERENCES

Almor, A., and S. Sloman, 1996. Is deontic reasoning special? *Psychol. Rev.* 103:374–380.

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

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

Barrett, H. C., 1999. *Guilty minds: How perceived intent, incentive, and ability to cheat influence social contract reasoning.* Paper presented at the 11th Annual Meeting of the Human Behavior and Evolution Society, Salt Lake City, Utah.

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

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

Cheng, P., and K. Holyoak, 1985. Pragmatic reasoning schemas. *Cogn. Psychol.* 17:391–416.

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

Cosmides, L., 1985. *Deduction or Darwinian algorithms? An explanation of the "elusive" content effect on the Wason selection task.* Doctoral diss., 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., H. C. Barrett, and J. Tooby, 2004. Social contracts elicit the detection of intentional cheaters, not innocent mistakes. Unpublished manuscript.

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

Cosmides, L., and J. Tooby, 1992. Cognitive adaptations for social exchange. In *The Adapted Mind*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press, pp. 163–228.

Cosmides, L., and J. Tooby, 1997. Dissecting the computational architecture of social inference mechanisms. In *Characterizing Human Psychological Adaptations Ciba Found. Symp.* 208:132–156.

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

Dugatkin, L. A., 1997. *Cooperation Among Animals: A Modern Perspective.* New York, Oxford University Press.

Fiddick, L., 1998. *The Deal and the Danger: An Evolutionary Analysis of Deontic Reasoning.* Doctoral diss., Department of Psychology, University of California, Santa Barbara.

Fiddick, L., 2004. Domains of deontic reasoning: Resolving the discrepancy between the cognitive and moral reasoning literatures. *Q. J. Exp. Psychol.* 57(3):447–474.

Fiddick, L., L. Cosmides, and J. Tooby, 2000. No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. *Cognition* 77:1–79.

Fiske, A., 1991. *Structures of Social Life: The Four Elementary Forms of Human Relations.* New York: Free Press.

Fodor, J., 2000. Why we are so good at catching cheaters. *Cognition* 75:29–32.

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

Gurven, M., 2002. To give and to give not: The behavioral ecology of human food transfers. *Behav. Brain Sci.*

Harris, P., and M. Núñez, 1996. Understanding of permission rules by preschool children. *Child Dev.* 67:1572–1591.

Harris, P., M. Núñez, and C. Brett, 2001. Let's swap: Early understanding of social exchange by British and Nepali children. *Mem. Cogn.* 29:757–764.

Hauser, M. D., in press. *Ought! The Inevitability of a Universal Moral Grammar.* New York: Holt.

Hasegawa, T., and K. Hiraishi, 2000. Ontogeny of the mind from an evolutionary psychological viewpoint. In *Comparative Cognitive Science of Mind,* S. Watanabe, ed. (in Japanese). Kyoto: Minerva Publications, pp. 413–427.

Isaac, G., 1978. The food-sharing behavior of protohuman hominids. *Sci. Am.* 238:90–108.

Johnson-Laird, P., and R. Byrne, 1991. *Deduction.* Hillsdale, N.J.: Erlsbaum.

Kirby, K., 1994. Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition* 51:1–28.

Maljkovic, V., 1987. *Reasoning in evolutionarily important domains and schizophrenia: Dissociation between content-dependent and content independent reasoning.* Thesis, Department of Psychology, Harvard University.

Manktelow, K. I., and J. St. B. T. Evans, 1979. Facilation of reasoning by realism: Effect or non-effect? *Br. J. Psychol.* 70: 477–488.

Manktelow, K., and D. Over, 1987. Reasoning and rationality. *Mind Lang.* 2:199–219.

Manktelow, K., and D. Over, 1988. *Sentences, stories, scenarios, and the selection task.* Paper presented at the First International Conference on Thinking, Plymouth, U.K.

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

Manktelow, K., and D. Over, 1991. Social roles and utilities in reasoning with deontic conditionals. *Cognition* 39:85–105.

Maynard Smith, J., 1982. *Evolution and the Theory of Games.* Cambridge, England: Cambridge University Press.

McKenna, P., L. Clare, and A. Baddeley, 1995. Schizophrenia. In *Handbook of Memory Disorders,* A. D. Baddeley, B. A. Wilson, and F. N. Watts, eds. New York: Wiley.

Núñez, M., and P. Harris, 1998a. Psychological and deontic concepts: Separate domains or intimate connections? *Mind Lang.* 13:153–170.

Núñez, M., and P. Harris, 1998b. *Young children's reasoning about prescriptive rules: Spotting transgressions through the selection task.* Paper presented at the XVth Biennial Meeting of the International Society for the Study of Behavioral Development, Berne, Switzerland.

Oaksford, M., and N. Chater, 1994. A rational analysis of the selection task as optimal data selection. *Psychol. Rev.* 101:608–631.

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

Platt, R., and R. Griggs, 1993. Darwinian algorithms and the Wason selection task: A factorial analysis of social contract selection task problems. *Cognition* 48:163–192.

Rips, L., 1994. *The Psychology of Proof.* Cambridge, Mass.: MIT Press.

Shallice, T., 1988. *From Neuropsychology to Mental Structure.* Cambridge, England: Cambridge University Press.

Sperber, D., F. Cara, and V. Girotto, 1995. Relevance theory explains the selection task. *Cognition* 57:31–95.

Stone, V., L. Cosmides, J. Tooby, N. Kroll, and R. Knight, 2002. Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proc. Natl. Acad. Sci. U.S.A.* 99:11531–11536.

Sugiyama, L., J. Tooby, and L. Cosmides, 2002. Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proc. Natl. Acad. Sci. U.S.A.* 99:11537–11542.

Tooby, J., and L. Cosmides, 1989. The logic of threat. Paper presented at a meeting of the Human Behavior and Evolution Society, Evanston, Ill., August.

Tooby, J., and L. Cosmides, 1992. The psychological foundations of culture. In *The Adapted Mind,* J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press, pp. 19–136.

Tooby, J., and L. Cosmides, 1996. Friendship and the Banker's Paradox: Other pathways to the evolution of adaptations for altruism. In *Evolution of Social Behaviour Patterns in Primates and Man,* W. G. Runciman, J. Maynard Smith, and R. I. M. Dunbar, eds. *Proceedings of the British Academy,* vol. 88, pp. 119–143.

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

Wason, P., 1966. Reasoning. In *New Horizons in Psychology,* B. M. Foss, ed. Harmondsworth: Penguin.

Wason, P., 1983. Realism and rationality in the selection task. In *Thinking and Reasoning: Psychological Approaches,* J. St. B. T. Evans, ed. London: Routledge, pp. 44–75.

Wason, P., and P. Johnson-Laird, 1972. *The Psychology of Reasoning: Structure and content.* Cambridge, Mass.: Harvard University Press.

Williams, G., 1966. *Adaptation and Natural Selection.* Princeton, N.J.: Princeton University Press.