CHAPTER 20

Neurocognitive Adaptations Designed for Social Exchange

LEDA COSMIDES and JOHN TOOBY

If a person doesn't give something to me, I won't give anything to that person. If I'm sitting eating, and someone like that comes by, I say, "Uhn, uhn. I'm not going to give any of this to you. When you have food, the things you do with it make me unhappy. If you even once in a while gave me something nice, I would surely give some of this to you."

Nisa from *Nisa: The Life and Words of a !Kung Woman*, Shostak, 1981, p. 89

Instead of keeping things, [!Kung] use them as gifts to express generosity and friendly intent, and to put people under obligation to make return tokens of friendship. . . . In reciprocating, one does not give the same object back again but something of comparable value.

Eland fat is a very highly valued gift . . . Toma said that when he had eland fat to give, he took shrewd note of certain objects he might like to have and gave their owners especially generous gifts of fat.

Marshall, 1976, pp. 366–369

Nisa and Toma were hunter-gatherers, !Kung San people living in Botswana's inhospitable Kalahari desert during the 1960s. Their way of life was as different from that in an industrialized, economically developed society as any on earth, yet their sentiments are as familiar and easy to comprehend as those of your neighbor next door. They involve *social exchange,* interactions in which one party provides a benefit to the other conditional on the recipient's providing a benefit in return (Cosmides, 1985; Cosmides & Tooby, 1989; Tooby & Cosmides, 1996). Among humans, social exchange can be implicit or explicit, simultaneous or sequential, immediate or deferred, and may involve alternating actions by the two parties or follow more complex structures. In all these cases, however, it is a way people cooperate for mutual benefit. Explicitly agreed-to forms of social exchange are the focus of study in economics (and are known as exchange or trade), while biologists and anthropologists focus more on implicit, deferred cases of exchange, often called *reciprocal altruism* (Trivers, 1971), *reciprocity,* or *reciprocation.* We will refer to the inclusive set of cases of the mutually conditioned provisioning of benefits as social exchange, regardless of subtype. Nisa and Toma are musing about social exchange interactions in which the expectation of reciprocity is implicit and the favor can be returned at a much later date. In their society, as in ours, the benefits given and received need not be physical objects for exchange to exist, but can be services (valued actions) as well. Aid in a fight, support in a political conflict, help with a sick child, permission to hunt and use water holes in your family's territory—all are ways of doing or repaying a favor. Social exchange behavior is both panhuman and ancient. What cognitive abilities make it possible?

For 25 years, we have been investigating the hypothesis that the enduring presence of social exchange interactions among our ancestors has selected for cognitive mechanisms that are specialized for reasoning about social exchange. Just as a lock and key are designed to fit together to function, our claim is that the proprietary procedures and conceptual elements of the social exchange reasoning specializations evolved to reflect the abstract, evolutionarily recurring relationships present in social exchange interactions (Cosmides & Tooby, 1989).

We picked social exchange reasoning as an initial test case for exploring the empirical power of evolutionary psychological analysis for a number of reasons. First, the topic is intrinsically important: Exchange is central to all human economic activity. If exchange in our species is made possible by evolved, neurocomputational programs specialized for exchange itself, this is surely worth knowing. Such evolved programs would constitute the foundation of economic behavior, and their specific properties would organize exchange interactions in all human societies; thus, if they exist, they deserve to be mapped. The discovery and mapping of such mechanisms would ground economics in the evolutionary and cognitive sciences, cross-connecting economics to the rest of the natural sciences. Social exchange specializations (if they exist) also underlie many aspects of a far broader category of implicit social interaction lying outside economics, involving favors, friendship, and self-organizing cooperation.

There was a second reason for investigating the computational procedures engaged by social exchange: The underlying counterhypothesis about social exchange reasoning that we have been testing against is the single most central assumption of the traditional social and behavioral sciences—the blank slate view of the mind that lies at the center of what we have called the *standard social science model* (Tooby & Cosmides, 1992). On this view, humans are endowed with a powerful, general cognitive capacity (intelligence, rationality, learning, instrumental reasoning), which explains human thought and the great majority of human behavior. In this case, humans putatively engage in successful social exchange through exactly the same cognitive faculties that allow them to do everything else: Their general intelligence allows them to recognize, learn, or reason out intelligent, beneficial courses of action. Despite—or perhaps because—this hypothesis has been central to how most neural, psychological, and social scientists conceptualize human behavior, it is almost never subjected to potential empirical falsification (unlike theories central to physics or biology). Investigating reasoning about social exchange provided an opportunity to test the blank slate hypothesis empirically in domains (economics and social behavior) where it had previously been uncritically accepted by almost all traditional researchers. Moreover, the results of these tests would be powerfully telling for the general issue of whether an evolutionary psychological program would lead to far-reaching and fundamental revisions across the human sciences. Why? If mechanisms of general rationality exist and are to genuinely explain anything of significance, they should surely explain social exchange reasoning as one easy application. After all, social exchange is absurdly simple compared to other cognitive activities such as language or vision, it is mutually beneficial and intrinsically rewarding, it is economically rational (Simon, 1990), and it should emerge spontaneously as the result of the ability to pursue goals; even artificially intelligent agents capable of pursuing goals through means-ends analysis should be able to manage it. An organism that was in fact equipped with a powerful, general intelligence would not *need* cognitive specializations for social exchange to be able to engage in it. If it turns out that humans nonetheless have adaptive specializations for social exchange, it would imply that mechanisms of general intelligence (if they exist) are relatively weak, and natural selection has specialized a far larger number of comparable cognitive competences than cognitive and behavioral scientists had anticipated.

Third, we chose reasoning because reasoning is widely considered to be the quintessential case of a content-independent, general-purpose cognitive competence. Reasoning is also considered to be the most distinctively human cognitive ability—something that exists in opposition to, and as a replacement for, instinct. If, against all expectation, even human reasoning turned out to fractionate into a diverse collection of evolved, content-specialized procedures, then adaptive specializations are far more likely to be widespread and typical in the human psychological architecture, rather than nonexistent or exceptional. Reasoning presents the most difficult test case, and hence the most useful case to leapfrog the evolutionary debate into genuinely new territory. In contrast, the eventual outcome of debates over the evolutionary origins and organization of motivation (e.g., sexual desire) and emotion (e.g., fear) are not in doubt (despite the persistence of intensely fought rearguard actions by traditional research communities). No blank slate process could, even in principle, acquire the human complement of motivational and emotional organization (Cosmides & Tooby, 1987; Tooby, Cosmides, & Barrett, 2005). Reasoning will be the last redoubt of those who adhere to a blank slate approach to the human psychological architecture.

Fourth, logical reasoning is subject to precise formal computational analysis, so it is possible to derive exact and contrasting predictions from domain-general and domain-specific theories, allowing critical tests to be devised and theories to be potentially or actually falsified.

Finally, we chose the domain of social exchange because it offered the opportunity to explore whether the evolutionary dynamics newly charted by evolutionary game theory (e.g., Maynard Smith, 1982) could be shown empirically to have sculpted the human brain and mind and, indeed, human moral reasoning. If it could be empirically shown that the kinds of selection pressures modeled in evolutionary game theory had real consequences on the human psychological architecture, then this would help lay the foundations of an evolutionary approach to social psychology, social behavior, and morality (Cosmides & Tooby, 2004). Morality was considered by most social scientists (then as now) to be a cultural product free of biological organization. We thought on theoretical grounds there should be an evolved set of domain-specific grammars of moral and social reasoning (Cosmides & Tooby, 1989) and wanted to see if we could clearly establish at least one rich empirical example—a grammar of social exchange. One pleasing feature of the case of social exchange is that it can be clearly traced step by step as a causal chain from replicator dynamics and game theory to details of the computational architecture to specific patterns of reasoning performance to specific cultural phenomena, moral intuitions, and conceptual primitives in moral philosophy—showcasing the broad integrative power of an evolutionary psychological approach. This research is one component of a larger project that includes mapping the evolutionary psychology of moral sentiments and moral emotions alongside moral reasoning (e.g., Cosmides & Tooby, 2004; Lieberman, Tooby, & Cosmides, 2003; Price, Cosmides, & Tooby, 2002).

What follows are some of the high points of this 25-year research program. We argue that social exchange is ubiquitously woven through the fabric of human life in all human cultures everywhere, and has been taking place among our ancestors for millions and possibly tens of millions of years. This means social exchange interactions are an important and recurrent human activity with sufficient time depth to have selected for specialized neural adaptations. Evolutionary game theory shows that social exchange can evolve and persist only if the cognitive programs that cause it conform to a narrow and complex set of design specifications. The complex pattern of functional and neural dissociations that we discovered during a 25-year research program 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 to 4) and appears cross-culturally—hunter-horticulturalists in the Amazon detect cheaters as reliably as adults who live in advanced market economies. The detailed patterns of human reasoning performance elicited by situations involving social exchange correspond to the evolutionarily derived predictions of a specialized logic or grammar of social exchange and falsify content-independent, general-purpose reasoning mechanisms as a plausible explanation for reasoning in this domain. A developmental process that is itself specialized for social exchange appears to be responsible for building the neurocognitive specialization found in adults: As we show, the design, ontogenetic timetable, and cross-cultural distribution of social exchange are not consistent with any known domain-general learning process. Taken together, the data showing design specificity, precocious development, cross-cultural universality, and neural dissociability implicate the existence of an evolved, species-typical neurocomputational specialization.

In short, the neurocognitive system that causes reasoning about social exchange shows evidence of being what Pinker (1994) has called a *cognitive instinct:* It is complexly organized for solving a well-defined adaptive problem our ancestors faced in the past, it reliably develops in all normal humans, 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. We briefly review the evidence that supports this conclusion, along with the evidence that eliminates the alternative by-product hypotheses that have been proposed. (For more comprehensive treatments, see Cosmides, 1985, 1989;

Cosmides & Tooby, 1989, 1992, 2005; Fiddick, Cosmides, & Tooby, 2000; Stone, Cosmides, Tooby, Kroll, & Knight, 2002; Sugiyama, Tooby, & Cosmides, 2002.)

SOCIAL EXCHANGE IN ZOOLOGICAL AND CULTURAL PERSPECTIVE

Living in daily contact affords many opportunities to see when someone needs help, to monitor when someone fails to help but could have, and, as Nisa explains, to withdraw future help when this happens. Under these conditions, reciprocity can be delayed, understanding of obligations and entitlements can remain tacit, and aid (in addition to objects) can be given and received (Shostak, 1981). But when people do not live side by side, social exchange arrangements typically involve explicit agreements, simultaneous transfer of benefits, and increased trade of objects (rather than intimate acts of aid). Agreements are explicit because neither side can know the other's needs based on daily interaction, objects are traded because neither side is present to provide aid when the opportunity arises, and trades are simultaneous because this reduces the risk of nonreciprocation—neither side needs to trust the other to provide help in the future. Accordingly, explicit or simultaneous trade is usually a sign of social distance (Tooby & Cosmides, 1996). !Kung, for example, will trade hides for knives and other goods with Bantu people but not with fellow band members (Marshall, 1976).

Explicit trades and delayed, implicit reciprocation differ in these superficial ways, but they share a deep structure: *X* provides a benefit to *Y* conditional on *Y* doing something that *X* wants. As humans, we take it for granted that people can make each other better off than they were before by exchanging benefits—goods, services, acts of help and kindness. But when placed in zoological perspective, social exchange stands out as an unusual phenomenon whose existence requires explanation. 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). Indeed, uncontroversial examples of social exchange in other species are difficult to find, and despite widespread investigation, social exchange has been reported in only a tiny handful of other species, such as chimpanzees, certain monkeys, and vampire bats (see Dugatkin, 1997; Hauser, in press, for contrasting views of the nonhuman findings).

Practices can be widespread without being the specific product of evolved psychological adaptations. Is social exchange a recent cultural invention? Cultural inventions such as alphabetic writing systems, cereal cultivation, and Arabic numerals are widespread, but they have one or a few points of origin, spread by contact, and are highly elaborated in some cultures and absent in others. Social exchange does not fit this pattern. It is found in every documented culture past and present and is a feature of virtually every human life within each culture, taking on a multiplicity of elaborate forms, such as returning favors, sharing food, reciprocal gift giving, explicit trade, and extending acts of help with the implicit expectation that they will be reciprocated (Cashdan, 1989; Fiske, 1991; Gurven, 2002; Malinowski, 1922; Mauss, 1925/1967). Particular methods or institutions for engaging in exchange—marketplaces, stock exchanges, money, the Kula Ring—are recent cultural inventions, but not social exchange behavior itself.

Moreover, evidence supports the view that social exchange is at least as old as the genus *Homo* and possibly far older than that. Paleoanthropological evidence indicates that before anatomically modern humans evolved, hominids engaged in social exchange (see, e.g., Isaac, 1978). Moreover, the presence of reciprocity in chimpanzees (and even certain monkeys; Brosnan & de Waal, 2003; de Waal, 1989, 1997a, 1997b; de Waal & Luttrell, 1988) suggests it may predate the time, 5 to 7 million years ago, when the hominid line split from chimpanzees. In short, social exchange behavior has been present during the evolutionary history of our line for so long that selection could well have engineered complex cognitive mechanisms specialized for engaging in it.

Natural selection retains and discards properties from a species' design based on how well these properties solve adaptive problems—evolutionarily recurrent problems whose solution promotes reproduction. To have been a target of selection, a design had to produce beneficial effects, measured in reproductive terms, in the environments in which it evolved. Social exchange clearly produced beneficial effects for those who successfully engaged in it, ancestrally as well as now (Cashdan, 1989; Isaac, 1978). A life deprived of the benefits that reciprocal cooperation provides would be a Hobbesian nightmare of poverty and social isolation, punctuated by conflict. But the fact that social exchange produces beneficial effects is not sufficient for showing that the neurocognitive system that enables it was designed by natural selection for that function. To rule out the counter-hypothesis that social exchange is a side effect of a system that was designed to solve a different or more inclusive set of adaptive problems, we need to evaluate whether the adaptation shows evidence of special design for the proposed function (Williams, 1966).

So what, exactly, is the nature of the neurocognitive machinery that enables exchange, and how specialized is it for this function? Social exchange is zoologically rare, raising the possibility that natural selection engineered into the human brain information processing circuits that are narrowly specialized for understanding, reasoning about, motivating, and engaging in social exchange. On this view, the circuits involved are neurocognitive adaptations *for* social exchange, evolved cognitive instincts designed by natural selection for that function—the *adaptive specialization hypothesis.* An alternative family of theories derives from the possibility that our ability to reason about and engage in social exchange is a by-product of a neurocognitive system that evolved for a different function. This could be an alternative specific function (e.g., reasoning about obligations). More usually, however, researchers expect that social exchange reasoning is a by-product or expression of a neurocognitive system that evolved to perform a more general function—operant conditioning, logical reasoning, rational decision making, or some sort of general intelligence. We call this family of explanations the *general rationality hypothesis.*

The general rationality hypothesis is so compelling, so self-evident, and so entrenched in our scientific culture that researchers find it difficult to treat it as a scientific hypothesis at all, exempting it from demands of falsifiability, specification, formalization, consistency, and proof they would insist on for any other scientific hypothesis. For example, in dismissing the adaptive specialization hypothesis of social exchange without examining the evidence, Ehrlich (2002) considers it sufficient to advance the folk theory that people just "figure it out." He makes no predictions nor specifies any possible test that could falsify his view. Orr (2003) similarly refuses to engage the evidence, arguing that perhaps "it just pays to behave in a certain way, and an organism with a big-enough brain reasons this out, while evolved instincts and specialized mental modules are beside the point" (p. 18). He packages this argument with the usual and necessarily undocumented claims about the low scientific standards of evolutionary psychology (in this case, voiced by unnamed colleagues in molecular biology).

What is problematic about this debate is not that the general rationality hypothesis is advanced as an alternative explanation. It is a plausible (if hopelessly vague) hypothesis. Indeed, the entire social exchange research program has, from its inception, been designed to systematically test against the major predictions that can be derived from this family of countertheories, to the extent they can be specified. What is problematic is that critics engage in the pretense that tests of the hypothesis they favor have never been carried out; that their favored hypothesis has no empirical burden of its own to bear; and that merely stating the general rationality hypothesis is enough to establish the empirical weakness of the adaptive specialization hypothesis. It is, in reality, what Dawkins (1986) calls the *argument from personal incredulity* masquerading as its opposite—a commitment to high standards of hypothesis testing.

Of course, to a cognitive scientist, Orr's conjecture as stated does not rise to the level of a scientific hypothesis. "Big brains" cause reasoning only by virtue of the neurocognitive programs they contain. Had Orr specified a reasoning mechanism or a learning process, we could empirically test the proposition that it predicts the observed patterns of social exchange reasoning. But he did not. Fortunately, however, a number of cognitive scientists have proposed some well-formulated by-product hypotheses, all of which make different predictions from the adaptive specialization hypothesis. Moreover, even where well-specified theories are lacking, one can derive some general predictions from the class of general rationality theories about possible versus impossible patterns of cultural variation, the effects of familiarity, possible versus impossible patterns of neural dissociation, and so on. We have tested each by-product hypothesis in turn. None can explain the patterns of reasoning performance found, patterns that were previously unknown and predicted in advance by the hypothesis that humans have neurocognitive adaptations designed for social exchange.

SELECTION PRESSURES AND PREDICTED DESIGN FEATURES

To test whether a system is an adaptation that evolved for a particular function, one must produce design evidence. The first step is to demonstrate that the system's properties solve a well-specified adaptive problem in a well-engineered way (Tooby & Cosmides, 1992, Chapter 1, this volume; Dawkins, 1986; Williams, 1966). This requires a well-specified theory of the adaptive problem in question.

For example, the laws of optics constrain the properties of cameras and eyes: Certain engineering problems must be solved by any information processing system that uses reflected light to project images of objects onto a 2-D surface (film or retina). Once these problems are understood, the eye's design makes sense. The transparency of the cornea, the ability of the iris to constrict the pupillary opening, the shape of the lens, the existence of photoreactive molecules in the retina, the resolution of retinal cells—all are solutions to these problems (and have their counterparts in a camera). Optics constrain the design of the eye, but the design of programs causing social behavior is constrained by the behavior of other agents—more precisely, by the design of the behavior-regulating programs in other agents and the fitness consequences that result from the interactions these programs cause. These constraints can be analyzed using evolutionary game theory (Maynard Smith, 1982).

An *evolutionarily stable strategy* (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). The rules of reasoning and decision making that guide social exchange in humans would not exist unless they had outcompeted alternatives, so we should expect that they implement an ESS.[1] By using game theory and conducting computer simulations of the evolutionary process, one can determine which strategies for engaging in social exchange are ESSs.

Selection pressures favoring social exchange exist whenever one organism (the provider) can change the behavior of a target organism to the provider's advantage by making the target's receipt of that benefit *conditional* on the target acting in a required manner. 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 (i.e., *If-then*) 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). For example, Toma knew that people in his band recognize and implicitly follow a social contract rule: *If you accept a generous gift of eland fat from someone, then you must give that person something valuable in the future.* Nisa's words also express a social contract: *If you are to get food in the future from me, then you must be individual Y* (where *Y* = an individual who has willingly shared food with Nisa in the past). Both realize that the act of accepting a benefit from someone triggers an obligation to behave in a way that somehow benefits the provider, now or in the future.

This mutual provisioning of benefits, each conditional on the other's compliance, is usually modeled by game theorists as a repeated Prisoners' Dilemma (Trivers, 1971; Axelrod & Hamilton, 1981; Boyd, 1988; but see Stevens & Stephens, 2004; Tooby & 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 & Tooby, 1989). Here, we 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 on.

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.,

[1] If the rules regulating reasoning and decision-making about social exchange do not implement an ESS, it would imply that these rules are a by-product of some other adaptation that produces fitness benefits so huge that they compensate for the systematic fitness costs that result from its producing non-ESS forms of social exchange as a side effect. Given how much social exchange humans engage in, this alternative seems unlikely.

Axelrod, 1984; Axelrod & Hamilton, 1981; Boyd, 1988; Trivers, 1971; Williams, 1966).[2] 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 in return. 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. Unconditional helping is not an ESS.

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 reciprocate (cheater designs). Therefore, conditional helping, which requires the ability to detect cheaters, is an ESS.

Engineers always start with a task analysis before considering possible design solutions. We did, too. By applying ESS analyses to the behavioral ecology of hunter-gatherers, we were able to specify tasks that an information processing program would have to be good at solving for it to implement an evolutionarily stable form of social exchange (Cosmides, 1985; Cosmides & Tooby, 1989). This task analysis of the required computations, *social contract theory,* specifies what counts as good design in this domain.

Because social contract theory provides a standard of good design against which human performance can be measured, there can be a meaningful answer to the question, "Are the programs that cause reasoning about social exchange well engineered for the task?" Well-designed programs for engaging in social exchange—if such exist—should include features that execute the computational requirements specified by social contract theory, and do so reliably, precisely, and economically (Williams, 1966).

From social contract theory's task analyses, we derived a set of predictions about the design features that a neurocognitive system specialized for reasoning about social exchange should have (Cosmides, 1985; Cosmides & Tooby, 1989). The following six design features (D1-D6) were among those on the list:

[2] Detecting cheaters is necessary for contingent cooperation to evolve, even when providing a benefit is cost free (i.e., even for situations that do not fit the payoff structure of a Prisoners' Dilemma; Tooby & Cosmides, 1996). In such cases, a design that cooperates contingently needs to detect when someone has failed to provide a benefit because it needs to know when to shift partners. In this model (just as in the Prisoners' Dilemma), a design that cannot shift partners will have lower fitness than a design that detects cheaters and directs future cooperation to those who do not cheat. Fitness is lower because of the opportunity cost associated with staying, not because of the cost of providing a benefit to the partner. Failure to understand that social exchange is defined by contingent provision of benefits, not by the suffering of costs, has resulted in some irrelevant experiments and discussion in the psychological literature. For example, showing that cheater detection can still occur when the requirement is not costly (e.g., Cheng & Holyoak, 1989) is a prediction of social contract theory, not a refutation of it (Cosmides, 1985; Cosmides & Tooby, 1989). For the same reason, there is no basis in social contract theory for Cheng and Holyoak's (1989) distinction between "social exchanges" (in which satisfying the requirement involves transferring a good, at some cost) and "social contracts" (in which satisfying a requirement may be cost free). For further discussion, see Fiddick et al. (2000).

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 interpretive procedures should not categorize that rule as a social contract. To trigger the inferences about obligations and entitlements that are appropriate to social contracts, the rule must be interpreted as restricting access to a benefit to those who have met a requirement. (This is a necessary, but not sufficient, condition; Cosmides & Tooby, 1989; Gigerenzer & Hug, 1992.)

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 also depends on which agent's point of view is taken. Perspective matters because the item, action, or state of affairs 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: individuals equipped with programs that cheat by design.[3] Hence, intentional cheating should powerfully trigger the detection system whereas mistakes should trigger it weakly or not at all. (Mistakes that result in an individual being cheated are relevant only insofar as they may not be true mistakes.)

D5. The hypothesis that the ability to reason about social exchange is acquired through the operation of some general-purpose learning ability necessarily predicts that good performance should be a function of experience and familiarity. In contrast, an evolved system for social exchange should be designed to recognize and reason about social exchange interactions no matter how unfamiliar the interaction may be, provided it can be mapped onto the abstract structure of a social contract. Individuals need to be able to reason about each new exchange situation as it arises, so rules that fit the template of a social contract should elicit high levels of cheater detection, even if they are unfamiliar.

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 & Tooby, 1989).

[3] *Programs that cheat by design* is a more general formulation of the principle, which does not require the human ability to form mental representations of intentions or to infer the presence of intentional mental states in others. An analogy to deception may be useful: Birds that feign a broken wing to lure predators away from their nests are equipped with programs that are designed to deceive the predator, but the cognitive procedures involved need not include a mental representation of an *intention* to deceive.

Cheating does involve the violation of a conditional rule, but note that it is 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 blind; 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 shown later, the definition of cheating implied by design features D1 through D4 does not map onto this content-blind definition of violation. What counts as cheating in social exchange is so content sensitive 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 program 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 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 you to look for potential violations of a conditional rule of the form *If P, then Q* (Wason, 1966, 1983; Wason & Johnson-Laird, 1972). Using this task, an extensive series of experiments has been conducted that addresses the following questions:

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

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 20.1 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; e.g., 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* and *Q*. (The italicized *P*s and *Q*s are not in problems given to subjects.)

conditional rules. Yet studies with the Wason selection task show that they are not. Consider the Wason task in Figure 20.1. 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 answer is not obvious to people. Studies in many nations have shown that reasoning performance is low on descriptive (indicative) rules like the rule in Figure 20.1: Only 5% to 30% of people give the logically correct answer, even when the rule involves familiar terms drawn from everyday life (Cosmides, 1989; Wason, 1966, 1983; Manktelow & Evans, 1979; Sugiyama et al., 2002). Interestingly, explicit instruction in logical inference does not boost performance: People who have just completed a semester-long college course in logic perform no better than people without this formal training (Cheng, Holyoak, Nisbett, & Oliver, 1986).

Formal logics, such as the propositional calculus, provide a standard of good design for content-general conditional reasoning: Their inference rules were constructed by philosophers to generate true conclusions from true premises, regardless of the subject matter one is asked to reason about. When human performance is measured against this standard, there is little evidence of good design: Conditional rules with descriptive content fail to elicit logically correct performance from 70% to 95% of people. Therefore, one can reject the hypothesis that the human mind is equipped with cognitive machinery that causes good conditional reasoning across all content domains.

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 a social contract? 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 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*). These are the cases that represent potential cheaters (Figure 20.2a). 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% to 80% of subjects get it right, the highest performance found for a task of this kind (for reviews, see Cosmides, 1985, 1989; Cosmides & Tooby, 1992, 1997; Fiddick et al., 2000; Gigerenzer & Hug, 1992; Platt & Griggs, 1993).

Given the content-blind 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 with Ebbinghaus disease (*P*) and the person who is not forgetful (*not-Q*) for the Ebbinghaus problem in Figure 20.1. 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 conditional reasoning problems (Cheng & Holyoak, 1985; Cosmides, 1989; Hasegawa & Hiraishi, 2000; Platt & Griggs, 1993; Sugiyama et al., 2002; supports D5, D6). Their minds distinguish social exchange content from other domains, and reason as if they were translating their terms into representational primitives such as *benefit, cost, obligation, entitlement, intentional,* and *agent* (Figure 20.2b; Cosmides & Tooby, 1992; Fiddick et al., 2000). Reasoning problems could be sorted into indefinitely many categories

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 20.2 Wason Task with a Social Contract Rule. (A) In response to this social contract problem, 76% of subjects chose *P and 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 20.1. 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 20.4, switched social contracts).

Table 20.1
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).
B9.	That rational choice theory can explain social contract effects.
B10.	That statistical learning produces the mechanisms that cause social contract reasoning.

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 in advance of its discovery by social contract theory's adaptationist analysis.

This pattern of good performance on reasoning problems involving social exchange is what we would expect if the mind reliably develops neurocognitive adaptations for reasoning about social exchange. But more design evidence is needed. Later we review experiments conducted to test for design features D1 through D6: features that should be present if a system specialized for social exchange exists.

In addition to producing evidence of good design for social exchange, recall that one must also show that the system's properties are not better explained as a solution to an alternative adaptive problem or by chance (Tooby & Cosmides, 1992, Chapter 1, this volume). 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 by-product and design feature implications are discussed in tandem. As we 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 20.1 lists the by-product hypotheses that have been tested and eliminated.

DO UNFAMILIAR SOCIAL CONTRACTS ELICIT CHEATER DETECTION? (D5)

An individual 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.

The evidence supports social contract theory: Cheater detection occurs even when the social contract is wildly unfamiliar (Figure 20.3a). 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 an individual must satisfy to be eligible for that benefit). When given this context, this outlandish, culturally alien rule elicits the same high level of cheater detection as highly familiar social exchange rules. This surprising result has been replicated for many different unfamiliar rules (Cosmides, 1985, 1989; Cosmides & Tooby, 1992; Gigerenzer & Hug, 1992; Platt & Griggs, 1993).

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.

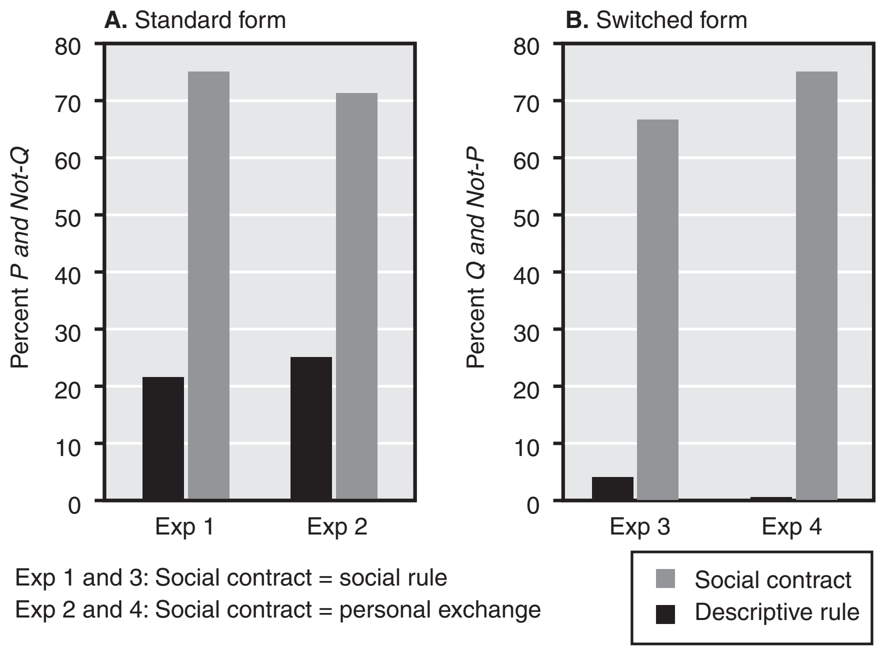

Figure 20.3 Detecting Violations of Unfamiliar Conditional Rules: Social Contracts versus Descriptive Rules. In these experiments, the same, unfamiliar rule was embedded either in a story that caused it to be interpreted as a social contract or in a story that caused it to be interpreted as a rule describing 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 and not-Q* also happens to be the logically correct answer. Over 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 rules (see Figure 20.4).

Familiarity is neither necessary nor sufficient for eliciting high performance (B1 of Table 20.1).

First, familiarity does not produce high levels of performance for descriptive rules (Cosmides, 1989; Manktelow & Evans, 1979). Note, for example, that the Ebbinghaus problem in Figure 20.1 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 for ones that are completely familiar (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)

As shown earlier, it is possible to construct social contract problems that will elicit a logically correct answer. But this is not because social exchange content activates logical reasoning.

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; see Figure 20.4). When given such problems, people look for cheaters, thereby giving a logically incorrect answer (*Q* and *not-P*).

PERSPECTIVE CHANGE

As predicted (D3), the mind's automatically deployed definition of cheating is tied to the perspective you are taking (Gigerenzer & 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 cheating employees. But it is not *logically* correct.[4]

In social exchange, the benefit to one agent is the requirement for the other: For example, giving pensions to employees benefits the employees but is the require-

[4] Moreover, the propositional calculus contains no rules of inference that allow *If P, then Q* to be translated as *If Q, then P* (i.e., no rule for translating [1] as [2]; see text) and then applying the logical definition of violation to [2] to arrive at the employee perspective answer (see Fiddick et al., 2000).

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 to whom this offer was made. 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 20.4 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). Check marks indicate the correct card choices if one is looking for cheaters—these 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* and *not-P*) when the rule is expressed in the switched format, and a logically correct answer (*P* and *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 20.3B). Note that a logically correct response to a switched social contract—where *P* = *requirement satisfied* and *not-Q* = *benefit not accepted*—would fail to detect cheaters.

ment the employer must satisfy (in exchange for > 10 years of employee service). To capture the distinction between the perspectives of the two agents, rules of inference for social exchange must be content sensitive, defining benefits and requirements relative to the agents involved. Because logical procedures are blind to the content of the propositions over which they operate, they have no way of representing the values of an action to each agent involved.

Switched Social Contracts

By moving the benefit from the antecedent clause (*P*) to the consequent clause (*Q*), one can construct a social exchange problem for which the adaptively correct cheater detection response is logically incorrect.

According to the propositional calculus (a formal logic), *If P then Q* does not imply *If Q then P*; therefore, "If you take the benefit, then you are obligated to satisfy the requirement," does not imply, "If you satisfy the requirement, then you are entitled to take the benefit." But inferential rules specialized for social exchange do license the latter inference (Cosmides & Tooby, 1989). Consequently, social exchange inferences (but not logical ones) should cause rule [1] above to be interpreted as implying:

> [2] If an employee has worked for the firm for over 10 years, then that employee gets a pension.

Assume you are concerned that employees have been cheating and are asked to check whether any employees have violated the rule. Although [2] and [1] are not logically equivalent, our minds interpret them as expressing the same social contract agreement. Hence, in both cases, a subroutine for detecting cheaters should cause you to check employees who have taken the benefit (gotten a pension) and employees who have not met the requirement (worked < 10 years).

But notice that these cards fall into different logical categories when the benefit to the potential cheater is in the antecedent clause versus the consequent clause (standard versus switched format, respectively; Figure 20.4). When the rule is expressed in the switched format, "got a pension" corresponds to the logical category *Q,* and "worked less than 10 years" corresponds to the logical category *not-P.* This answer will correctly detect employees who are cheating, but it is logically incorrect. When the rule is expressed in the standard format, the same two cards correspond to *P* and *not-Q.* For standard format social contracts, the cheater detection subroutine will produce the same answer as logical procedures would—not because this response is logically correct, but because it will detect cheaters.

When given switched social contracts like [2], subjects overwhelmingly respond by choosing *Q & not-P,* a logically incorrect answer that correctly detects cheaters (Figure 20.3b; Cosmides, 1985, 1989; Gigerenzer & Hug, 1992; supports D2, D6). Indeed, when subjects' choices are classified by *logical* category, it looks like standard and switched social contracts elicit different responses. But when their choices are classified by *social contract* category, they are invariant: For both rule formats, people choose the cards that represent an agent who took the benefit and an agent who did not meet the requirement.

This robust pattern occurs precisely because social exchange reasoning is sensitive to content: It responds to a syntax of agent-relative benefits and requirements, not antecedents and consequents. Logical procedures would fail to detect cheaters on switched social contracts. Being content blind, their inferential rules are doomed to checking *P* and *not-Q,* even when these cards correspond to potential altruists (or fools)—that is, to people who have fulfilled the requirement and people who have not accepted the benefit.

ELIMINATING LOGIC (B2, B3)

Consider the following by-product hypothesis: The dissociation between social contracts and descriptive rules is not caused by a cheater detection mechanism.

Instead, the human cognitive architecture applies content-free rules of logical inference, such as *modus ponens* and *modus tollens.* These logical rules are activated by social contract content but not by other kinds of content, and that causes the spike in *P & not-Q* answers for social contracts.

The results of the switched social contract and the perspective change experiments eliminate this hypothesis. Social contracts elicit a logically incorrect answer, *Q & not-P,* when this answer would correctly detect cheaters. Logical rules applied to the syntax of the material conditional cannot explain this pattern, because these rules would always choose a true antecedent and false consequent (*P & not-Q*), never a true consequent and false antecedent (*Q & not-P*).

There is an active debate about whether the human cognitive architecture includes content-blind rules of logical inference, which are sometimes dormant and sometimes activated (e.g., Bonatti, 1994; Rips, 1994; Sperber, Cara, & Girotto, 1995). We are agnostic about that issue. What is clear, however, is that such rules cannot explain reasoning about social contracts (for further evidence, see Fiddick et al., 2000).

DEDICATED SYSTEM OR GENERAL INTELLIGENCE?

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, & Baddeley, 1995). Yet their ability to detect cheaters can remain intact. Maljkovic (1987) tested the reasoning of patients suffering from positive symptoms of schizophrenia, comparing their performance with that of hospitalized (nonpsychotic) 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. Indeed, it was indistinguishable from 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, which can operate even when the systems responsible for more general reasoning are damaged. It provides further support for the claim that social exchange reasoning is functionally and neurally distinct from more general abilities to process information or behave intelligently.

HOW MANY SPECIALIZATIONS FOR CONDITIONAL REASONING?

Social contracts are not the only conditional rules for which natural selection should have designed specialized reasoning mechanisms (Cosmides, 1989). Indeed, good violation detection is also found for conditional rules drawn from two other domains: threats and precautions. Is good performance across these three domains caused by a single neurocognitive system or by several functionally distinct ones? If a single system causes reasoning about all three domains, then we should not claim that cheater detection is caused by adaptations that evolved for that specific function.

The notion of multiple adaptive specializations is commonplace in physiology: The body is composed of many organs, each designed for a different function. Yet many psychologists cringe at the notion of multiple adaptive specializations when these are computational. Indeed, evolutionary approaches to psychology foundered in the early 1920s on what was seen as an unfounded multiplication of "instincts."

That was before the cognitive revolution, with its language for describing what the brain does in information processing terms and its empirical methods for revealing the structure of representations and processes. Rather than relying on a priori arguments about what should or could be done by a single mechanism, we can now empirically test whether processing about two domains is accomplished by one mechanism or two. We should not imagine that there is a separate specialization for solving each and every adaptive problem. Nor should real differences in processing be ignored in a misguided effort to explain all performance by reference to a single mechanism. As Einstein once said, "Make everything as simple as possible, but no simpler."

Conditional Reasoning about Other Social Domains

Threats specify a conditional rule (*If you don't do what I require, I will harm you*), which the threatener can violate in two ways: by bluffing or by double-crossing. It appears that people are good at detecting bluffs and double-crosses on Wason tasks that test threats (with an interesting sex difference never found for social exchange problems; Tooby & Cosmides, 1989). However, these violations do not map onto the definition of cheating and, therefore, cannot be detected by a cheater detection mechanism. This suggests that reasoning about social contracts and threats is caused by two distinct mechanisms. (So far, no theory advocating a single mechanism for reasoning about these two domains has been proposed. Threats are not deontic; see later discussion.)

Also of adaptive importance is the ability to detect when someone is in danger by virtue of having violated a precautionary rule. These rules have the general form, "*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 wear a gas mask"). Using the Wason task, it has been shown that people are very good at detecting potential violators of precautionary rules; that is, individuals who have engaged in a hazardous activity without taking the appropriate precaution (e.g., those working with toxic gases [*P*] and those not wearing a gas mask [*not-Q*]). Indeed, relative to descriptive rules, precautions show a spike in performance, and the magnitude of this content effect is about the same as that for detecting cheaters on social contracts (Cheng & Holyoak, 1989; Fiddick et al., 2000; Manktelow & Over, 1988, 1990, 1991; Stone et al., 2002).

A system well designed for reasoning about hazards and precautions should have properties different from one for detecting cheaters, many of which have been tested for and found (Fiddick, 1998, 2004; Fiddick et al., 2000; Pereyra & Nieto, in press; Stone et al., 2002). Therefore, alongside a specialization for reasoning about social exchange, the human cognitive architecture should contain computational machinery specialized for managing hazards, which causes good violation detection on precautionary rules. Obsessive-compulsive disorder, with its compulsive worrying, checking, and precaution taking, may be caused by a misfiring of this precautionary system (Cosmides & Tooby, 1999; Leckman & Mayes, 1998, 1999).

An alternative view is that reasoning about social contracts and precautionary rules is generated by a single mechanism. Some view both social contracts and precautions as deontic rules (i.e., rules specifying obligations and entitlements) and wonder whether there is a general system for reasoning about deontic conditionals. More specifically, Cheng and Holyoak (1985, 1989) have proposed that inferences about both types of rule are generated by a permission schema, which operates over a larger class of problems.[5]

Can positing a permission schema explain the full set of relevant results? Or are they more parsimoniously explained by positing two separate adaptive specializations, one for social contracts and one for precautionary rules? We are looking for a model that is as simple as possible, but no simpler.

SOCIAL CONTRACT ALGORITHMS OR A PERMISSION SCHEMA? LOOKING FOR DISSOCIATIONS *WITHIN* THE CLASS OF PERMISSION RULES (D1, D2, D4)

Permission rules are a species of conditional rule. According to Cheng and Holyoak (1985, 1989), these rules are imposed by an authority to achieve a social purpose, and they specify the conditions under which an individual is permitted to take an action. Cheng and Holyoak speculate that repeated encounters with such social rules cause domain-general learning mechanisms to induce a *permission schema,* consisting of four production rules (see Table 20.2 on p. 606). This schema generates inferences about any conditional rule that fits the following template: "If action A is to be taken, then precondition R must be satisfied."

Social contracts fit this template. In social exchange, an agent *permits* you to take a benefit from him or her, conditional on your having met the agent's requirement. There are, however, many situations other than social exchange in which an action is permitted conditionally. Permission schema theory predicts uniformly high performance for the entire class of permission rules, a set that is larger, more general, and more inclusive than the set of all social contracts (see Figure 20.5 on p. 607).

On this view, a neurocognitive system specialized for reasoning about social exchange, with a subroutine for cheater detection, does not exist. According to their hypothesis, a permission schema causes good violation detection for all permission rules; social contracts are a subset of the class of permission rules; therefore, cheater detection occurs as a by-product of the more domain-general permission schema (Cheng & Holyoak, 1985, 1989).

In contrast, the adaptive specialization hypothesis holds that the design of the reasoning system that causes cheater detection is more precise and functionally specialized than the design of the permission schema. Social contract algorithms should have design features that are lacking from the permission schema, such as responsivity to benefits and intentionality. As a result, removing benefits (D1, D2) and/or intentionality (D4) from a social contract should produce a permission rule that fails to elicit good violation detection on the Wason task.

[5] Cheng and Holyoak (1985) also propose an obligation schema, but permission and obligation schemas do not lead to different predictions on the kinds of rules usually tested (see Cosmides, 1989; Rips, 1994, p. 413).

Table 20.2
The Permission Schema Is Composed of Four Production Rules[a]

Rule 1:	If the action is to be taken, then the precondition must be satisfied.[b]
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.

[a] Cheng and Holyoak, 1985.
[b] 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.

As Sherlock Holmes might put it, we are looking for the dog that did not bark: permission rules that do *not* elicit good violation detection. That discovery would falsify permission schema theory. Social contract theory predicts functional dissociations *within* the class of permission rules whereas permission schema theory does not.

NO BENEFITS, NO SOCIAL EXCHANGE REASONING: TESTING D1 AND D2

To trigger cheater detection (D2) and inference procedures specialized for interpreting social exchanges (D1), a rule needs to regulate access to benefits, not to actions more generally. Does reasoning performance change when benefits are removed?

Benefits Are Necessary for Cheater Detection (D1, D2)

The function of a social exchange for each participant is to gain access to benefits 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. *Taking a benefit* is a representational primitive within the social contract template *If you take benefit B, then you must satisfy requirement R.*

The permission schema template has representational primitives with a larger scope than that proposed for social contract algorithms. For example, *taking a benefit* is *taking an action*, but not all cases of taking actions are cases of taking benefits. As a result, all social contracts are permission rules, but not all permission rules are social contracts. Precautionary rules can also be construed as permission rules (although they need not be; see Fiddick et al., 2000, exp. 2). They, too, have a more restricted scope: *Hazardous actions* are a subset of *actions; precautions* are a subset of *preconditions.*

Note, however, that there are permission rules that are neither social contracts nor precautionary rules (see Figure 20.5). This is because there are actions an individual can take that are not *benefits* (social contract theory) and that are not *hazardous* (hazard management theory). Indeed, we encounter many rules like this in everyday life—bureaucratic and corporate rules, for example, often state a procedure that is to be followed without specifying a benefit (or a danger). If the mind has a permission schema, then people should be good at detecting violations of

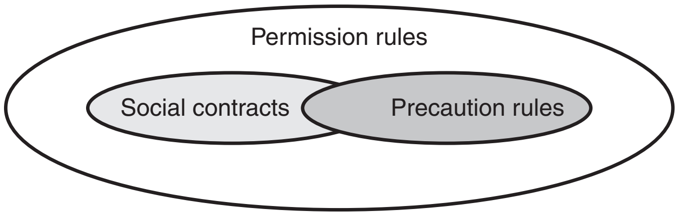

Figure 20.5 The Class of Permission Rules Is Larger Than, and Includes, Social Contracts and Precautionary Rules. Many of the permission rules we encounter in everyday life are neither social contracts nor precautions (white area). Rules of civil society (etiquette, customs, traditions), bureaucratic 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 20.2) 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.

rules that fall into the white area of Figure 20.5, that is, permission rules that are neither social contracts nor precautionary. But they are not. Benefits are necessary for cheater detection.

Using the Wason task, several abs have tested permission rules that involve no benefit (and are not precautionary). As predicted by social contract theory, these do not elicit high levels of violation detection. For example, Cosmides and Tooby (1992) constructed Wason tasks in which the elders (authorities) were creating laws governing the conditions under which adolescents are permitted to take certain actions. For all tasks, the law fit the template for a permission rule. The permission rules tested differed in just one respect: whether the action to be taken is a benefit or an unpleasant chore. The critical conditions compared performance on these two rules:

> [3] "If one goes out at night, then one must tie a small piece of red volcanic rock around one's ankle."
>
> [4] "If one takes out the garbage at night, then one must tie a small piece of red volcanic rock around one's ankle."

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 (getting to go out at night), 80% of subjects answered correctly; when it was a chore (taking out the garbage), only 44% did so. This dramatic decrease in violation detection was predicted in advance by social contract theory. Moreover, it violates the central prediction of permission schema theory: that being a permission rule is sufficient to facilitate violation detection. There are now many experiments showing poor violation detection with permission rules that lack a benefit (e.g., Barrett, 1999; Cosmides, 1989, exp. 5; Fiddick, 2003; Manktelow & Over, 1991; Platt & Griggs, 1993).

This is another dissociation by content, but this time it is *within* the domain of permission rules. To elicit cheater detection, a permission rule must be interpreted as restricting access *to a benefit.* It supports the psychological reality of the representational primitives posited by social contract theory, showing that the representations necessary to trigger differential reasoning are more content specific than those of the permission schema.

Benefits Trigger Social Contract Interpretations (D1)

The Wason experiments just described tested D1 and D2 in tandem. But D1—the claim that benefits are necessary for permission rules to be *interpreted* as social contracts—receives support independent of experiments testing D2 from studies of moral reasoning. Fiddick (2004) asked subjects what justifies various permission rules and when an individual should be allowed to break them. The rules were closely matched for surface content, and context was used to vary their interpretation. The permission rule that lacked a benefit (a precautionary one) elicited different judgments from permission rules that restricted access to a benefit (the social contracts). Whereas social agreement and morality, rather than facts, were more often cited as justifying the social contract rules, facts (about poisons and antidotes) rather than social agreement were seen as justifying the precautionary rule. Whereas most subjects thought it was acceptable to break the social contract rules if you were not a member of the group that created them, they thought the precautionary rule should always be followed by people everywhere. Moreover, the explicit exchange rule triggered very specific inferences about the conditions under which it could be broken: Those who had received a benefit could be released from their obligation to reciprocate, *but only by those who had provided the benefit to them* (i.e., the obligation could not be voided by a group leader or by a consensus of the recipients themselves). The inferences subjects made about the rules restricting access to a benefit follow directly from the grammar of social exchange laid out in social contract theory (Cosmides & Tooby, 1989). These inferences were not—and should not—be applied to precautionary rules (see also Fiddick et al., 2000). The presence of a benefit also predicts inferences about emotional reactions to seeing someone violate a permission rule: Social contract violations were thought to trigger anger whereas precautionary violations were thought to trigger fear (Fiddick, 2004). None of these dissociations within the realm of permission rules are predicted by permission schema theory.

INTENTIONAL VIOLATIONS VERSUS INNOCENT MISTAKES: TESTING D4

Intentionality plays no role in permission schema theory. Whenever the action has been taken but the precondition has not been satisfied, the permission schema should register that a *violation* has occurred. As a result, people should be good at detecting violations of permission rules, whether the violations occurred by accident or by intention. In contrast, social contract theory predicts a mechanism that looks for *intentional* violations (D4).

Program designs that cause unconditional helping are not evolutionarily stable strategies. Conditional helping can be an ESS because cheater detection provides a specific fitness advantage unavailable to unconditional helpers: By identifying cheaters, the conditional helper can avoid squandering costly cooperative efforts in the future on those who, by virtue of having an alternative program design, will not reciprocate. This means the evolutionary function of a cheater detection subroutine is to correctly connect an attributed disposition (to cheat) with a person (a cheater). It is not simply to recognize instances wherein an individual did not get what he or she was entitled to. 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, does not reveal the disposition or design of the exchange partner. Accidents may result in someone being cheated, but without indicating the presence of a cheater.[6]

Therefore, social contract theory predicts an additional level of cognitive specialization beyond looking for violations of a social contract. Accidental violations of social contracts will not fully engage the cheater detection subroutine; intentional violations will (D4).

A Dissociation for Social Contracts

Given the same social exchange 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 inattentive but well meaning to a condition 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, & Tooby, forthcoming; supports D4; disconfirms B1-B8). Fiddick (1998, 2004) found the same effect (as did Gigerenzer & Hug, 1992, using a different context manipulation).

In both scenarios, violating the rule would result in someone being cheated, yet high performance occurred only when being cheated was caused by a cheater. 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 (D4) or by the violator's failure to benefit from her mistake (D2; see earlier discussion, on the necessity of *benefits* to elicit cheater detection). He found that both factors independently contributed to the drop, equally and additively. Thus, the same decrease in performance occurred whether (1) violators would benefit from their innocent mistakes, or (2) violators wanted to break the rule on purpose but would not benefit from doing so. For scenarios missing both factors (i.e., accidental violations that do not benefit the violator), performance dropped by twice as much as when just one factor was missing. That is, the more factors relevant to cheater detection are removed, the more performance dropped.

In bargaining games, experimental economists have found that subjects are twice as likely to punish defections (failures to reciprocate) when it is clear that the defector intended to cheat as when the defector is a novice who might have simply made a mistake (Hoffman, McCabe, & Smith, 1998). This provides

[6] Mistakes can be faked, of course. Too many by a given individual should raise suspicion, as should a single mistake that results in a very large benefit. Although this prediction has not been tested yet, we would expect social contract algorithms to be sensitive to these conditions.

interesting convergent evidence, using entirely different methods, for the claim that programs causing social exchange distinguish between mistakes and intentional cheating.

NO DISSOCIATION FOR PRECAUTIONS

Different results are expected for precautionary rules. Intentionality should not matter if the mechanisms that detect violations of precautionary rules were designed to look for people in danger. For example, a person who is not wearing a gas mask while working with toxic gases is in danger, whether that person forgot the gas mask at home (accidental violation) or left it home on purpose (intentional violation). That is, varying the intentionality of a violation should affect social exchange reasoning but not precautionary reasoning. Fiddick (1998, 2004) tested and confirmed this prediction: Precautionary rules elicited high levels of violation detection whether the violations were accidental or intentional, but performance on social contracts was lower for accidental violations than for intentional ones. This functional distinction between precautionary and social exchange reasoning was predicted in advance based on the divergent adaptive functions proposed for these two systems.

ELIMINATING PERMISSION SCHEMA THEORY (B4)

The preceding results violate central predictions of permission schema theory. According to that theory, (1) all permission rules should elicit high levels of violation detection, whether the permitted action 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 permitted action is neutral or unpleasant (yet not hazardous). Moreover, people are bad at detecting accidental violations of permission rules that are social contracts. Taken together, these results eliminate the hypothesis that the mind contains or develops a permission schema of the kind postulated by Cheng and Holyoak (1985, 1989).

ELIMINATING CONTENT-FREE DEONTIC LOGICS (B6)

The same results also falsify hypothesis B6: that cheater detection on social contracts is caused by a content-free deontic logic (for discussion of this possibility, see Manktelow & Over, 1987). All the benefit and intentionality tests described in this section involved deontic rules, but not all elicited high levels of violation detection.

This same set of results also defeats a related claim by Fodor (2000): that "the putative cheater detection effect on the Wason task is actually a materials artifact" (p. 29). This sweeping conclusion is predicated on the (mistaken) notion that the only evidence for cheater detection comes from experiments in which the control problems are indicative (i.e., descriptive) conditional rules (a curious mistake because it is refuted by experiments with deontic controls, which are presented in the single source Fodor cites: Cosmides & Tooby, 1992). According to Fodor, reasoning *from* a deontic conditional rule that is stipulated to hold is more likely to elicit violation detection than reasoning *about* a rule whose truth is in question (even though in both cases the individual is asked to do the same thing: look for rule violations). Fodor's explanation for this purported difference is deeply flawed (among other things, it assumes what it seeks to explain). But instead of disputing Fodor's reasoning, let us consider whether his artifact explanation can account for the cheater detection results observed. After all, there are many experiments comparing reasoning on social contracts to reasoning about other deontic conditionals.

According to Fodor, high levels of violation detection will be found for any deontic rule that specifies what people are (conditionally) required to do (because all involve reasoning with the law of contradiction). All the permission rules described earlier had precisely this property, all were stipulated to hold, and, in every case, subjects were asked to reason *from* the rule, not about it. If Fodor's artifact hypothesis were correct, all of these rules should have elicited good violation detection. But they did not. Violation detection was poor when the deontic rule lacked a benefit; it was also poor for social contract rules when the potential violator was accused of making innocent mistakes rather than intentional cheating. This pattern is predicted by social contract algorithms, but not by Fodor's hypothesis that reasoning from a deontic conditional rule is sufficient to elicit good violation detection.

B5—that social contract rules elicit good performance merely because we understand what implications follow from them (e.g., Almor & Sloman, 1996)—is eliminated by the intention versus accident dissociation. The same social contract rule—with the same implications—was used in both conditions. If the rule's implications were understood in the intention condition, they should also have been understood in the accident condition. Yet the accident condition failed to elicit good violation detection. Understanding the implications of a social contract may be necessary for cheater detection (Fiddick et al., 2000), but the accident results show this is not sufficient.

In short, it is not enough to admit that moral reasoning, social reasoning, or deontic reasoning is special: The specificity of design for social exchange is far narrower in scope.

A NEUROPSYCHOLOGICAL DISSOCIATION BETWEEN SOCIAL CONTRACTS AND PRECAUTIONS

Like social contracts, precautionary rules are conditional, deontic, and involve subjective utilities. 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, to deontic rules, or to deontic rules involving subjective utilities; Cheng & Holyoak, 1989; Manktelow & Over, 1988, 1990, 1991; Sperber et al., 1995). Most of these one-mechanism theories are undermined by the series of very precise, functional dissociations between social exchange reasoning and reasoning about other deontic permission rules (discussed earlier). But a very strong test, one that addresses *all* one-mechanism theories, would be to find a neural dissociation between social exchange and precautionary reasoning.

One Mechanism or Two?

If reasoning about social contracts and precautions is caused by a single mechanism, then neurological damage to that mechanism should lower performance on both types of rule. 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 et al. (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: percentage correct for precautions minus percentage correct for social contracts. For normal subjects, these difference scores were all close to zero (Mean = 1.2 percentage points, $SD = 11.5$).

Stone et al. (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. R. M.'s difference score (precautions minus social contracts) was 31 percentage points. This is 2.7 standard deviations larger than the average difference score of 1.2 percentage points found for control subjects ($p < .005$). In other words, R. M. had a large deficit in his social contract reasoning, alongside normal reasoning about precautionary rules.

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 licenses inferences about the underlying mental structures. 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 damage to a circuit connecting anterior temporal cortex to the amygdalae was important in creating R. M.'s selective deficit.[7] Recent functional imaging (fMRI) studies also support the hypothesis that social contract reasoning is supported by different brain areas than precautionary reasoning, and imply the involvement of several brain areas in addition to temporal cortex (Wegener, Baare, Hede, Ramsoy, & Lund, 2004; Fiddick, Spampinato, & Grafman, forthcoming).

[7] Stone et al. (2002) tested two other patients with overlapping but different patterns of brain damage. R.B. had more extensive bilateral orbitofrontal damage than R. M., and had some anterior temporal damage as well, 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.

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 neurocognitive system. R. M.'s dissociation is inconsistent with all of these one-mechanism accounts. These accounts include mental logic (Rips, 1994), mental models (Johnson-Laird & Byrne, 1991), decision theory/optimal data selection (Kirby, 1994; Oaksford & Chater, 1994), permission schema theory (Cheng & Holyoak, 1989), relevance theory (Sperber et al., 1995),[8] and Manktelow and Over's (1991, 1995) view implicating a system that is general to any deontic rule that involves subjective utilities. (For further evidence against relevance theory, see Fiddick et al., 2000; for further evidence against Manktelow & Over's theory, see Fiddick & Rutherford, in press.)

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-sensitive cognitive specialization within the human reasoning system.

PRECOCIOUS DEVELOPMENT OF SOCIAL EXCHANGE REASONING

Children understand what counts as cheating on a social contract by age 3 (Harris & Núñez, 1996; Harris, Núñez, & Brett, 2001; Núñez & Harris, 1998a).[9] This has been shown repeatedly in experiments by Harris and Núñez using an evaluation task: a task in which the child must decide when a character is violating a rule. Consider, for example, a story in which Carol wants to ride her bicycle but her mom says, "If you ride your bike, then you must wear an apron." This rule restricts access to a benefit (riding the bike) based on whether the child has satisfied an arbitrary requirement. The child is then shown four pictures (Carol riding the bike wearing an apron, Carol riding without an apron, Carol wearing an apron but not riding, and Carol not riding or wearing an apron) and asked to choose the picture in which Carol is doing something naughty. British 3-year-olds chose the correct picture (Carol riding the bike with no apron) 72% to 83% of the time; 4-year-olds, 77% to 100% of the time (Harris & Núñez, 1996; Harris et al., 2001; Núñez & Harris, 1998a). These performance levels were found whether the social contract emanated from the mother or was a consensual swap between two children; that is, the rule did not have to be imposed by an authority figure. A variety of tests showed that, for social contracts, children understood that taking the benefit was *conditional* on meeting the requirement. They were not merely looking for cases in which the requirement was not met; they were looking for cases in which the benefit was taken *and* the requirement was not met. The same effects were found for preschoolers from the United Kingdom, Colombia, and (with minor qualifications) rural Nepal.

The performance of the preschoolers was adultlike in other ways. Like adults, the preschoolers did well whether the social contract was familiar or unfamiliar.

[8] For a full account of the problems relevance theory has explaining social contract reasoning, see Fiddick et al., 2000.

[9] Younger children have not been tested yet.

Also like adults, intentionality mattered to the children. Núñez and Harris (1998a) varied (1) whether the character had taken the benefit or not and (2) whether the character had failed to fulfill the requirement by accident or deliberately. Children were far more likely to say the character had been naughty when the breach was intentional than accidental. Four-year-olds deemed social contract violations naughty 81% of the time when they were intentional versus 10% of the time when they were accidental; for 3-year-olds, the figures were 65% versus 17%, respectively. Children also could match emotions to outcomes for reciprocal exchanges: Given an agreement to swap, they understood that the victim of cheating would feel upset, and that both children would be happy if the swap was completed (Núñez, 1999).

Moreover, the children tested by Harris and Núñez (1996) showed the same dissociation between social contract and descriptive rules as adults: 3- to 4-year-olds chose the correct violation condition only 40% of the time for descriptive rules but 72% to 83% of the time for social contracts. By age 5, children could solve a full-array Wason selection task when the rule was a social contract (Núñez & Harris, 1998b; performance limitations, rather than competence problems, interfered with the Wason performance of the preschoolers).[10]

CROSS-CULTURAL INVARIANCES AND DISSOCIATIONS IN SOCIAL EXCHANGE REASONING

Cognitive neuroscientists have long been aware that neural dissociations are useful for elucidating mental structure. But cultural dissociations may provide a uniquely informative source of converging evidence. Because the ontogenetic experience of people in different cultures varies widely, cross-cultural studies allow one to see whether differences in ontogenetic experience are associated with differences in mental structure.

Most psychologists and anthropologists believe that high-level cognitive competences emerge from general-purpose cognitive abilities trained by culturally specific activities, rather than as part of our evolved, reliably developing, species-typical design. That cheater detection should be well developed across cultures is a falsifiable prediction of the evolutionary account, which posits that this competence should be distributed in a species-typical, human universal fashion. More precisely, because detecting cheaters is necessary for social exchange to be an ESS, the development of cheater detection should be buffered against cultural variation and, therefore, be uniform. In contrast, the development of ESS-irrelevant aspects of performance (e.g., interest in acts of generosity) is under no selection to be uniform across cultures and should, therefore, be free to vary with cultural circumstance.

[10] Although the definitive experiments have not yet been done, existing evidence suggests that preschoolers also understand violations of precautionary rules. The rules used by Harris and Núñez (1996) fell into two categories: pure social contracts ("arbitrary permissions" and "swaps," in their terminology) and hybrid rules (ones that can be interpreted either as social contracts or precautionary). The hybrids were rules that restricted access to a benefit on the condition that a precaution was taken, for example, *If you play outside, you must wear a coat* (to keep warm). Cummins (1996) tested a more purely precautionary rule, but the context still involved restrictions on access to a benefit (playing outside).

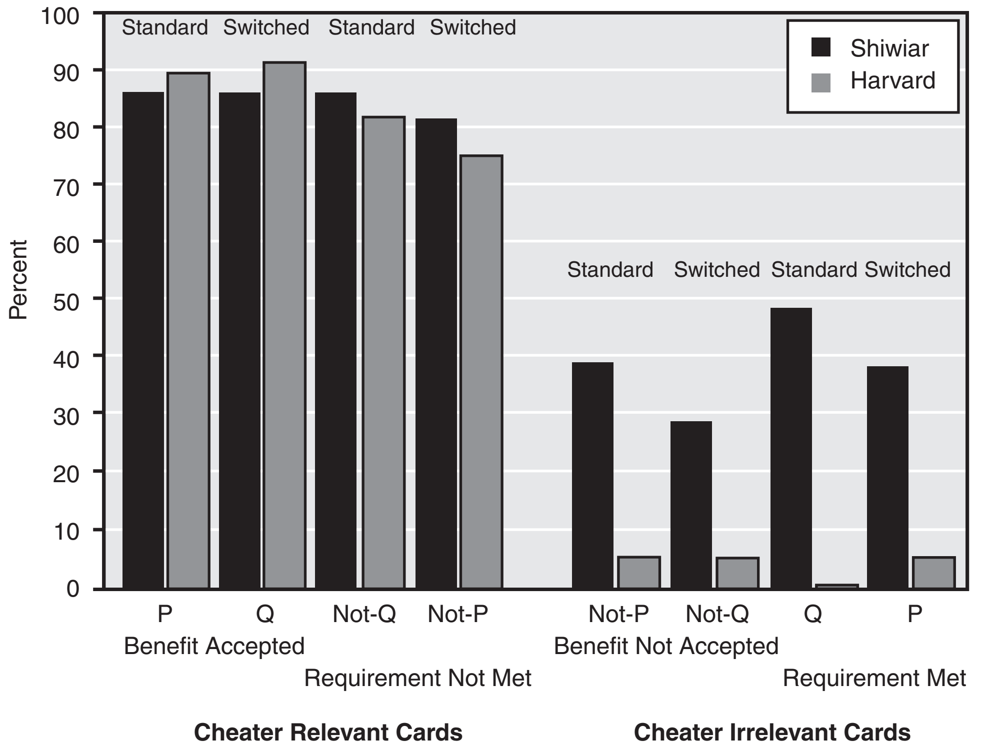

Figure 20.6 Performance of Shiwiar Hunter-Horticulturalists and Harvard Undergraduates on Standard and Switched Social Contracts. (Percent of subjects choosing each card.) There was no difference between the two populations in their choice of cheater relevant cards (*benefit accepted, requirement not satisfied*). They differed only in their choice of cheater-irrelevant cards (Shiwiar showing more 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. (E.g., the *P* card is cheater relevant for a standard social contract, but not for a switched one; see Figure 20.4.)

Sugiyama, Tooby, and Cosmides (2002) tested these predictions 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 facilitates cheating; markets increase the volume of transactions experienced by each individual. If no evolved specialization is involved—that is, if general-purpose processes induce a cheater detection subroutine through repeated experience with cheating—then this subroutine might not be found outside the Western world.

The Shiwiar were raised and continue to live in a culture as different from that of American college students as any on the planet. Nevertheless, Shiwiar were just as good at detecting cheaters on Wason tasks as Harvard undergraduates were (Figure 20.6). For cheater-relevant cards, the performance of Shiwiar hunter-horticulturalists was identical to that of Harvard students. Shiwiar differed only in that they were more likely to also show interest in cheater-irrelevant cards—the ones that could reveal acts of generosity. (Their excellence at cheater detection did not result from indiscriminate interest in all cards. Controlling for logical category, Shiwiar were more than twice as likely to choose a card when it was cheater-relevant than when it was not; $p < .005$.) In short, 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.

Is cheater detection invariant because the sociocultural experience of Shiwiar and American subjects is too similar to cause differences in reasoning performance? Clearly not; if that were true, the two populations would perform identically on cheater-irrelevant cards as well as on cheater-relevant ones. That did not happen.

This is the only research we know of to show identical performance across very different cultural groups on those aspects of a reasoning problem that are relevant to a cognitive adaptation functioning as an evolutionarily stable strategy, yet different performance on those aspects that are irrelevant to the adaptation functioning as an ESS. That performance in detecting cheaters was invariant across very disparate cultural settings suggests that the brain mechanism responsible is a reliably developing neurocognitive system. That is, its development is canalized in a way that buffers it against idiosyncratic variations in ontogenetic experience.

DOES DOMAIN-GENERAL LEARNING BUILD THE SPECIALIZATION FOR SOCIAL EXCHANGE?

The empirical evidence reviewed earlier strongly supports the claim that reasoning about social exchange is caused by neurocognitive machinery that is specialized for this function in adults: social contract algorithms. This conclusion was supported not just by evidence from Wason tasks but also from experimental economics games, moral reasoning protocols, emotion attribution tasks, and developmental studies. What makes the Wason results particularly interesting, however, is that the Wason task requires information search. The Wason results indicate the presence of a subroutine that is narrowly specialized for *seeking out* information that would *reveal* the presence of cheaters. This subroutine is not designed to seek out information that would reveal the presence of cheating (when this occurs by mistake), or permission violations, or violations in general.

But how was this very precisely designed computational specialization produced? Are the developmental mechanisms that build social contract algorithms domain-specific and specialized for this function? Or are social contract specializations in adults built by domain-general learning mechanisms?

If computational specializations for social exchange are acquired via some general-purpose learning process, then we should not claim that the specialization is an evolved adaptation *for* social exchange. Instead, the social exchange specialization would be the product of a learning mechanism that evolved to solve a different, perhaps more general, adaptive problem.

General Purpose Learning Is a Nonstarter

Evidence of an adaptive specialization in the adult human mind often meets the following rejoinder: Although the adult mechanism is specialized, the mechanisms that built it are not—the adult specialization was acquired via a general purpose learning process (e.g., Elman et al., 1996; Rumelhart & McClelland, 1986; Gauthier & Tarr, 2002; Orr, 2003; for discussion, see Duchaine, 2001; Pinker, 2002; Tooby & Cosmides, 1992).

There is a fundamental problem with this view: No general purpose learning process is known to science (Gallistel, 2000). This is not because scientists are in the dark about animal learning. Learning processes specialized for solving specific adaptive problems have been found in many species, including dead reckoning in desert ants, learned food aversions in rats, star navigation in birds, snake fear in primates, and language acquisition in humans (Gallistel, 1990, 2000; Garcia, 1990; Garcia & Koelling, 1966; Mineka & Cook, 1993; Pinker, 1994). Indeed, even classical conditioning, considered by many to be the premier example of general purpose learning, is anything but (Staddon, 1988). The empirical evidence shows that this form of learning is adaptively specialized for a specific computational task common in foraging and predator avoidance: multivariate nonstationary time series analysis (Gallistel & Gibbon, 2000).

Classical and operant conditioning are adaptive specializations, but it is true that they operate over inputs from many different domains (i.e., they are somewhat content-general). So let us reframe the rejoinder thus: Are adult specializations for reasoning about social exchange acquired via classical or operant conditioning?

At the root of operant and classical conditioning is the ability to respond contingently to reward and punishment (Gallistel & Gibbon, 2000; Staddon, 1988). Social exchange entails such contingencies: 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.

Yet the ability to respond contingently to reward and punishment is not sufficient for social exchange to emerge in a species. All animal species can be classically and operantly conditioned (Staddon, 1988), but few species engage in social exchange. If classical and/or operant conditioning caused the acquisition of social exchange specializations, then social exchange should be zoologically widespread. The fact that it is so rare means that it is not the consequence of any behavior-regulation or learning process that is zoologically common.

Although reciprocity is rare in the animal kingdom, it is found in a number of nonhuman primate species (Brosnan & de Waal, 2003; de Waal & Luttrell, 1988; de Waal, 1989, 1997a, 1997b). Its presence in other primates means that social exchange behavior can arise in the absence of language. This means the conditioning hypothesis cannot be rescued by arguing that the development of social exchange requires the joint presence of language and conditioning mechanisms.

Not Rational Choice (B9)

Can the development of neurocognitive specializations for reasoning about social exchange be accounted for by the fact that reciprocity is economically advantageous? An economic folk theory exists and was recently articulated by Orr (2003, p. 18):

> An evolutionary psychologist might counter that the fact that a behavior conforms so closely to what's expected of an adaptive one is evidence that it's a bona fide biological adaptation. And here we arrive at another problem. For the same logic that makes a behavior evolutionarily advantageous might also make it "economically" advantageous. . . . The point is that when evolutionary and economic considerations yield the same prediction, conformity to Darwinian predictions cannot be taken as decisive.

This would be a good point if economists had a theory of the computations that give rise to economic learning and decision making. But they do not. Having no account of how economic reasoning is accomplished, economists rely on rational choice theory, an *as if* approach. According to rational choice theory, people reason *as if* they were equipped with neurocognitive mechanisms that compute (in some as yet unspecified way) the subjective expected utility of alternative actions, and choose the one that maximizes personal utility (Savage, 1954).

Rational choice theory makes very precise predictions about the choices people should make when engaging in social exchange and other economic games. Contrary to Orr's assumption, however, rational choice theory and the evolutionarily functional theory of social exchange make different predictions about human behavior (Hoffman, McCabe, & Smith, 1998). There is now a large body of results from experimental economics showing that people rarely behave as rational choice theory predicts and that this is not due to inexperience with the experimental situation—even experienced subjects violate rational choice theory predictions (e.g., Fehr & Gächter, 2000a, 2000b; Henrich et al., in press; Hoffman, McCabe, & Smith, 1998). For example, when given the opportunity to engage in social exchange, people routinely and systematically choose to cooperate with others when they would earn a higher payoff by defecting; they also punish acts of cheating when they would earn more by not doing so. That is, they cooperate and punish in circumstances, such as the one-shot Prisoners' Dilemma, where these choices are not utility maximizing (Hoffman, McCabe, & Smith, 1998). As Hoffman, McCabe, and Smith (1998) argue, these are precisely the responses one would expect of specializations designed to operate in small hunter-gatherer bands, where repeated interactions are the norm and one-shot interactions are rare. The results reported earlier on accidental versus intentional violations of social contracts are also inconsistent with economic prediction. Rational choice theory predicts mechanisms that respond to the payoff structure of situations, not to intentions, and cheating produces the same negative payoff whether it was accidental or intentional. Thus, a system designed for maximizing utility should detect cheating, not cheaters. Yet that is not the empirical finding.

Rational or *economically advantageous* has to refer to some kind of reasoning process if it is to serve as an explanation of anything, and the most completely axiomatized normative model of rational economic behavior fails to predict or explain the facts of when humans choose to cooperate and punish, either in social exchange (Hoffman, McCabe, & Smith, 1998) or in public goods games (Fehr & Gächter, 2000a, 2000b; Henrich et al., in press; Kurzban, McCabe, Smith, & Wilson, 2001). Because the facts of social exchange reasoning and behavior contradict central predictions of rational choice theory, this economic by-product hypothesis cannot explain the features of the neurocognitive specialization found in adults, or the development of these features (B9 eliminated). In light of this failure, a number of economists are turning to evolutionary psychological accounts of social exchange and judgment under uncertainty to explain human economic behavior (Gigerenzer & Selten, 2001; Hoffman, McCabe, & Smith, 2001; Romer, 2000).

Statistical Learning and Content-Free Inductive Inference: More Dogs That Do Not Bark (B10)

Various accounts of inductive learning have been proposed: Bayesian learning machines, connectionist systems that compute a multiple regression, contingency calculators. Some posit highly domain-specific, inductive learning systems (e.g., Marcus, 2001; Staddon, 1988), but most do not (e.g., Elman et al., 1996; Quartz & Sejnowski, 1997).

The domain-general proposals foreground the role of content-blind inductive inference procedures in the construction of mental content.[11] These extract statistical relationships from patterns that are objectively present in the external world. Indeed, they are constrained to do so: The world is the only source of content for these statistical learning mechanisms. As a result, we should see certain dogs barking. For example, twentieth-century Chicago schoolchildren should fear things that are dangerous to children living in twentieth-century urban Chicago—electric sockets, cars, streets, hot stoves. The content of their fears should reflect the frequency and statistical distribution of dangers in the modern world because it was constructed by content-free mechanisms operating on information derived from these distributions.

By contrast, domain-specific learning mechanisms are content rich: They allow inferences that go beyond the information given, so the mental content constructed may be richer than (or merely different from) the statistical distribution of information in the external world of individual experience. For example, when asked what they are most afraid of, Chicago schoolchildren name lions, tigers, wild animals, "monsters" (dangerous but unspecified animal or humanlike creatures), snakes, and spiders (Maurer, 1965). The content of their fears reflects the statistical distribution of dangers in an ancestral world they have never experienced (Marks, 1987). It does not reflect the statistical distribution of dangers in urban Chicago—that is, the modern dogs are not barking.

People reliably develop—apparently by age 3—social contract algorithms with the properties discussed in this review. These properties make that neurocognitive system very good at solving an adaptive problem of the ancestral world: seeking out information that would reveal cheaters. We know there is good design for this ancestral problem because very precise patterns of dissociations by content—both functional and neural—were predicted in advance of their discovery on the basis of ESS analyses applied to the behavioral ecology of hunter-gatherers. However, statistical learning theories cannot even retrodict this pattern of dissociations (let alone predict them in advance).

The explanatory variables that drive statistical learning are experience, repetition, and their consequence, familiarity. If these variables caused the development

[11] Attentional biases (e.g., for faces) play a role in some of the domain-general theories (e.g., Elman et al., 1996), but these are thought to be few in number and, crucially, to not contain the mental content that is eventually constructed (the source of which is patterns in the world).

of reasoning specializations, we should observe a different set of reasoning specializations than are found, including ones that produce good violation detection for permission rules and even descriptive ones. But these modern dogs are not barking.

Where Is the Specialization for Finding Violations of Descriptive Rules? Descriptive rules are not rare, exotic occurrences. They are claims about how the world works, commonplaces of everyday conversation (*If you wait until November, the clinic will be out of flu shots. If she eats hot chili, she likes a cold beer. If you use that pan, the casserole will stick. If you wash with bleach, your clothes will be whiter.*). Actions are more likely to succeed when they are based on true rather than false information, so violations of these claims should be salient. Consistent with this, people do know what counts as a violation: They can tell you that cases in which *P* happens but *Q* does not violate a descriptive rule, even when the rule is abstract or unfamiliar (Manktelow & Over, 1987).

But this knowledge does not translate into efficacious information search. Although people *recognize* violations of descriptive rules when they occur, they do not *seek out* information that could reveal such violations, even when they are explicitly asked to do so on a Wason task (see instructions for Figure 20.1; for discussion, see Fiddick et al., 2000). That is, humans do not reliably develop reasoning specializations that cause them to *look for* potential violations of descriptive rules. This dissociation between people's knowledge and what information they search for is found for descriptive rules but not for social contracts. Descriptive rules are ubiquitous. If experience with a type of rule were sufficient for statistical learning to build a specialization for information search, then we should observe good violation detection on Wason tasks using descriptive rules (even unfamiliar ones), just as we do for social contracts.

Even worse, experience with *specific* descriptive rules does nothing to improve performance. Early research using the Wason task explored whether violation detection for descriptive rules was better when the rule, relation, or any of its terms were familiar. It was not (Cosmides, 1985; Cheng, Holyoak, Nisbett, & Oliver, 1986; Manktelow & Evans, 1979; Wason, 1983). Furthermore, people who had repeated experience with instances that violated a particular concrete rule performed no better than people who did not have these experiences (Manktelow & Evans, 1979). The impotence of repeated experience with concrete violations is mirrored in the social contract results, where high performance is observed regardless of experience. College students are intimately familiar with rules restricting access to alcohol (e.g., *If you drink beer, then you must be over 21*), yet Cosmides (1985) found they are no better at detecting violations of this familiar rule than they are for never-experienced rules about cassava root and tattoos.

Where Is the Specialization for Finding Violations of Permission Rules? The failure of statistical learning theories becomes even clearer when we consider that social exchange rules are but a small subset of all permission rules (which are, in turn, a subset of deontic rules, which are themselves a subset of all conditional rules). By class inclusion, humans necessarily have far more experience with permission rules than with social contracts (legend, Figure 20.5). 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, and that this schema should operate not only on social contracts but also on precautionary rules and indeed on any social norm that gives conditional permission. Yet careful tests showed that the permission schema they predicted does not exist.

Poor performance in detecting violations of conditional permission rules drawn from the white zone of Figure 20.5 cannot be explained by claiming that all the permission rules we happen to encounter are either social contracts or precautions. Conditional social norms that fit neither category permeate our society (*If one eats red meat, then one drinks red wine. If you live east of Milpas Street, then vote at Cleveland Elementary School. If the blue inventory form is filled out, file it in the metal bin.*). Yet we do not develop information search strategies specialized for detecting violations of such rules.

Where Is the Specialization for Detecting Negative Payoffs? Statistical learning theorists might respond by saying that learning occurs in response to negative payoffs (see Manktelow & Over, 1995, for a related proposal). This view predicts an information search specialization for detecting when a negative payoff might occur, whether it is produced by cheating on a social contract or failing to take precautions in hazardous situations (Manktelow & Over, 1991, 1995).

Fiddick and Rutherford (in press) show that no such specialization exists: Information search on Wason tasks using social contracts and related rules bears no relationship to subjects' judgments about which outcomes produce negative payoffs. Moreover, R. M.'s neural dissociation (preserved search for violations of precautionary rules with impaired search for cheaters) shows that the mind does not contain a unitary specialization for detecting negative payoffs.

Where Is the Specialization for Detecting Cheating, Rather than Cheaters? What if statistical learning is triggered by negative payoffs, but only within the domain of social exchange? (This is hardly a domain-general proposal, but never mind.) A person can be cheated—receive a negative payoff due to the violation of a social exchange agreement—by accident or by intention. Both kinds of violation damage personal utility, both are useful to detect, and both require detection if the participant in an exchange is to get what he or she wants and is entitled to. Moreover, because innocent mistakes and intentional cheating both result in someone being cheated, situations in which a person *was cheated* are statistically more common than situations in which someone was cheated *by a cheater.* Hence, this domain-restricted version of statistical learning predicts the development of an information search specialization that looks for acts in which someone was cheated, regardless of cause. This specialization would be easy to engineer: A mechanism that indiscriminately scrutinizes cases in which the benefit was accepted and cases in which the requirement was not met would reveal both accidental and intentional violations. But this specialization does not exist: People are not good at detecting acts of cheating when there is evidence that they occurred by accident rather than intention.

In contrast, it is specifically the detection of intentional cheaters that makes contingent exchange evolutionarily stable against exploitation by cheaters (i.e., an ESS). That people are good at detecting intentional cheating but not accidental mistakes is a unique prediction of the evolutionary task analysis of exchange.

Variables That Affect Statistical Learning Do Not Seem to Affect the Development of Cheater Detection An information search specialization for detecting cheaters reliably develops across large variations in experience, repetition, and familiarity. For example:

- Precocious performance is neither necessary nor sufficient for sustaining an adaptationist hypothesis (Cosmides & Tooby, 1997). It is, however, relevant for evaluating claims of content-free inductive learning because these predict that the development of reasoning skills will reflect the child's experience (e.g., Markman, 1989). The early age at which children understand social exchange reasoning undermines the hypothesis that social contract specializations were constructed by content-independent procedures operating on individual experience.

  Preschool-age children are not noted for the accuracy and consistency of their reasoning in many domains, even ones with which they have considerable experience. For example, many children this age will say that a raccoon can change into a skunk; that there are more daisies than flowers; that the amount of liquid changes when poured from a short fat beaker into a tall thin one; that they have a sister but their sister does not (Boden, 1980; Carey, 1984; Keil, 1989; Piaget, 1950). When reasoning about social exchange, however, preschool-age children show virtually all the features of special design that adults do.

  When a child has had experience in a number of domains, it is difficult to explain how or why a content-blind statistical learning mechanism would cause the early and uniform acquisition of a reasoning skill for one of these domains, yet fail to do so for the others. When one considers that adults have massive experience with permission rules, yet fail to develop specializations for detecting violations of this more general and, therefore, more common class, the presence of accurate cheater detection in 3- and 4-year-olds is even more surprising.
- Cultural experience 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.

Statistical Learning Summary Neither experience, repetition, nor familiarity explain which reasoning skills develop and which do not, yet they should if specializations develop via statistical learning. 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 the mechanisms responsible operate smoothly regardless of experience and familiarity; that they detect cheaters and not other kinds of violators; and that the developmental process results in a social contract specialization rather than one for more inclusive classes such as permission rules.

CONCLUSIONS

There are strict standards of evidence for claiming that an organic system is an evolved adaptation. The system that causes reasoning about social exchange meets these standards. Reasoning about social exchange narrowly dissociates from other forms of reasoning, both cognitively and neurally. The pattern of results reveals a system equipped with exactly those computational properties necessary to produce an evolutionarily stable form of conditional helping (as opposed to the many kinds of unconditional helping that are culturally encouraged). These properties include, but are not limited to, the six design features discussed herein, all of which were predicted in advance from the task analyses contained in social contract theory (see Cosmides & Tooby, 1992, Fiddick, Cosmides, & Tooby, 2000 for others). Importantly, the pattern of results cannot be explained as a by-product of a reasoning adaptation designed for some different, or more general, function. Every by-product hypothesis proposed in the literature has been tested and eliminated as an explanation for social exchange reasoning (see Table 20.1).

The design of the computational specialization that causes social exchange reasoning in adults (and preschoolers) places limits on any theory purporting to account for its development. No known domain-general process can account for the fact that social contract specializations with these particular design features reliably develop across cultures, whereas specializations for more commonly encountered reasoning problems do not develop at all. Indeed, the social contract specialization has properties that are better adapted to the small-group living conditions of ancestral hunter-gatherers than to modern industrial societies. Experience of the world may well be necessary for its development during ontogeny, but the developmental process implicated appears to be a domain-specific one, designed by natural selection to produce an evolutionarily stable strategy for conditional helping.

The simplest, most parsimonious explanation that can account for all the results—developmental, neuropsychological, cognitive, and behavioral—is that the human brain contains a neurocognitive adaptation designed for reasoning about social exchange. Because the developmental process that builds it is specialized for doing so, this neurocognitive specialization for social exchange reliably develops across striking variations in cultural experience. It is one component of a complex and universal human nature.

REFERENCES

Almor, A., & Sloman, S. (1996). Is deontic reasoning special? *Psychological Review, 103,* 374–380.

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

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

Barrett, H. C. (1999, June). *Guilty minds: How perceived intent, incentive, and ability to cheat influence social contract reasoning.* 11th annual meeting of the Human Behavior and Evolution Society, Salt Lake City, Utah.

Boden, M. (1980). *Jean Piaget.* New York: Viking.

Bonatti, L. (1994). Why should we abandon the mental logic hypothesis? *Cognition, 50,* 17–39.

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

Brosnan, S. F., & de Waal, F. B. M. (2003). Monkeys reject unequal pay. *Nature, 425,* 297–299.

Carey, S. (1984). Cognitive development: The descriptive problem. In M. S. Gazzaniga (Ed.), *Handbook of cognitive neuroscience.* (pp. 37–66). New York: Plenum Press.

Cashdan, E. (1989). Hunters and gatherers: Economic behavior in bands. In S. Plattner (Ed.), *Economic anthropology* (pp. 21–48). Stanford, CA: 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 (UMI No. 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., Barrett, H. C., & Tooby, J. (forthcoming). *Social contracts elicit the detection of intentional cheaters, not innocent mistakes.*

Cosmides, L., & Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupre (Ed.), *The latest on the best: Essays on evolution and optimality.* Cambridge, MA: MIT Press.

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

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 163–228). New York: Oxford University Press.

Cosmides, L., & Tooby, J. (1997). Dissecting the computational architecture of social inference mechanisms. In *Characterizing human psychological adaptations (Ciba Foundation Symposium #208)* (pp. 132–156). Chichester, England: Wiley.

Cosmides, L., & Tooby, J. (1999). Towards an evoluationary taxonomy of treatable conditions. *Journal of Abnormal Psychology, 108,* 453–464.

Cosmides, L., & Tooby, J. (2004). Knowing thyself: The evolutionary psychology of moral reasoning and moral sentiments. In R. E. Freeman & P. Werhane (Eds.), *Business, science, and ethics: The Ruffin series* (No. 4, pp. 93–128). Charlottesville, VA: Society for Business Ethics.

Cosmides, L., & Tooby, J. (2005). Social exchange: The evolutionary design of a neurocognitive system. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences, III.* Cambridge, MA: MIT Press.

Cummins, D. D. (1996). Evidence of deontic reasoning in 3- and 4-year-old children. *Memory and Cognition, 24,* 823–829.

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

de Waal, F. (1989). Food sharing and reciprocal obligations among chimpanzees. *Journal of Human Evolution, 18,* 433–459.

de Waal, F. (1997a). The chimpanzee's service economy: Food for grooming. *Evolution and Human Behavior, 18,* 375–386.

de Waal, F. (1997b). Food transfers through mesh in brown capuchins. *Journal of Comparative Psychology, 111,* 370–378.

de Waal, F., & Luttrell, L. (1988). Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethology and Sociobiology, 9,* 101–118.

Duchaine, B. (2001). *Computational and developmental specificity in face recognition: Testing the alternative explanations in a developmental prosopagnosic.* Doctoral dissertation, Department of Psychology, University of California, Santa Barbara.

Dugatkin, L. A. (1997). *Cooperation among animals: A modern perspective.* New York: Oxford.

Ehrlich, P. (2002). *Human natures: Genes, cultures, and the human prospect.* Washington, DC: Island Press.

Elman, J., Bates, E., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: Connectionism in a developmental context.* Cambridge, MA: MIT Press.

Fehr, E., & Gächter, S. (2000a). Cooperation and punishment in public goods experiments. *American Economic Review, 90,* 980–994.

Fehr, E., & Gächter, S. (2000b). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives, 14,* 159–181.

Fiddick, L. (1998). *The deal and the danger: An evolutionary analysis of deontic reasoning.* Doctoral dissertation, Department of Psychology, University of California, Santa Barbara.

Fiddick, L. (2003). Is there a faculty of deontic reasoning? A critical re-evaluation of abstract deontic versions of the Wason selection task. In D. Over (Ed.), *Evolution and the psychology of thinking: The debate* (pp. 33–60). Hove, England: Psychology Press.

Fiddick, L. (2004). Domains of deontic reasoning: Resolving the discrepancy between the cognitive and moral reasoning literatures. *Quarterly Journal of Experimental Psychology, 57A*(4), 447–474.

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

Fiddick, L., & Rutherford, M. (in press). Looking for loss in all the wrong places: Loss-avoidance does not explain cheater detection. *Evolution and Human Behavior.*

Fiddick, L., Spampinato, M., & Grafman, J. (forthcoming). Social contracts and precautions activate different neurological systems: An fMRI investigation of deontic reasoning.

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.

Gallistel, C. R. (1990). *The organization of learning.* Cambridge, MA: MIT Press.

Gallistel, C. R. (2000). The replacement of general-purpose learning models with adaptively specialized learning modules. In M. S. Gazzaniga (Eds.), *The New Cognitive Neurosciences* (pp. 1179–1191). Cambridge, MA: MIT Press.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate and conditioning. *Psychological Review, 107,* 289–344.

Garcia, J. (1990). Learning without memory. *Journal of Cognitive Neuroscience, 2*(4), 287–305.

Garcia, J., & Koelling, R. A. (1966). Relations of cue to consequence in avoidance learning. *Psychonomic Science, 4,* 123–124.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 431–446.

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

Gigerenzer, G., & Selton, R. (Eds.). (2001). *Bounded rationality: The adaptive toolbox.* Cambridge, MA: The MIT Press.

Gurven, M. (in press). To give and to give not: The behavioral ecology of human food transfers. *Behavioral and Brain Sciences,* ID code: bbs00001257.

Harris, P., & Núñez, M. (1996). Understanding of permission rules by preschool children. *Child Development, 67,* 1572–1591.

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

Hasegawa, T., & Hiraishi, K. (2000). Ontogeny of the mind from an evolutionary psychological viewpoint. In S. Watanabe (Ed.), *Comparative cognitive science of mind* (pp. 413–427). Kyoto: Minerva Publications.

Hauser, M. D. (in press). *Ought! The inevitability of a universal moral grammar.* New York: Holt.

Henrich, J., Boyd, R., Bowles, S., Gintis, H., Fehr, E., Camerer, C., et al. (in press). 'Economic man' in cross-cultural perspective. *Behavioral and Brain Sciences.*

Hoffman, E., McCabe, K., & Smith, V. (1998). Behavioral foundations of reciprocity: Experimental economics and evolutionary psychology. *Economic Inquiry, 36,* 335–352.

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

Johnson-Laird, P., & Byrne, R. (1991). *Deduction.* Hillsdale, NJ: Erlbaum.

Keil, F. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

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

Kurzban, R., McCabe, K., Smith, V. L., & Wilson, B. J. (2001). Incremental commitment and reciprocity in a real time public goods game. *Personality and Social Psychology Bulletin, 27*(12), 1662–1673.

Leckman, J., & Mayes, L. (1998). Maladies of love: An evolutionary perspective on some forms of obsessive-compulsive disorder. In D. M. Hann, L. C. Huffman, I. I. Lederhendler, & D. Meinecke (Eds.), *Advancing research on developmental plasticity: Integrating the behavioral science and neuroscience of mental health* (pp. 134–152). Rockville, MD: U.S. Department of Health and Human Services.

Leckman, J., & Mayes, L. (1999). Preoccupations and behaviors associated with romantic and parental love: Perspectives on the origin of obsessive-compulsive disorder. *Obsessive-Compulsive Disorder, 8*(3), 635–665.

Lieberman, D., Tooby, J., & Cosmides, L. (2003). Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceedings of the Royal Society London. (Biological Sciences), 270*(1517), 819–826.

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

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

Manktelow, K., & Evans, J. J., St. B. T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology, 70,* 477–488.

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

Manktelow, K., & Over, D. (1988, July). Sentences, stories, scenarios, and the selection task. *First international conference on thinking.* Plymouth, UK.

Manktelow, K., & Over, D. (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, pp. 153–164). London: Wiley.

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

Manktelow, K., & Over, D. (1995). Deontic reasoning. In S. E. Newstead & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honor of Peter Wason* (pp. 91–114). Hove, England: Erlbaum.

Marcus, G. (2001). *The algebraic mind: Reflections on connectionism and cognitive science.* Cambridge, MA: MIT Press.

Markman, E. (1989). *Categorization and naming in children.* Cambridge, MA: MIT Press.

Marks, I. (1987). *Fears, phobias, and rituals.* New York: Oxford University Press.

Marshall, L. (1976). Sharing, talking, and giving: Relief of social tensions among the !Kung. In R. Lee & I. DeVore (Eds.), *Kalahari hunter-gatherers: Studies of the !Kung San and their neighbors.* Cambridge, MA: Harvard University Press.

Maurer, A. (1965). What children fear. *Journal of Genetic Psychology, 106,* 265–277.

Mauss, M. (1967). *The gift: Forms and functions of exchange in archaic societies.* New York: Norton. (Original work published 1925)

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

McKenna, P., Clare, L., & Baddeley, A. (1995). Schizophrenia. In A. D. Baddeley, B. A. Wilson, & F. N. Watts (Eds.), *Handbook of memory disorders* (pp. 271–292). New York: Wiley.

Mineka, S., & Cook, M. (1993). Mechanisms involved in the observational conditioning of fear. *Journal of Experimental Psychology: General, 122,* 23–38.

Núñez, M. (1999). Conditional agreements of reciprocal exchange and the emotions involved. *IX European Conference on Developmental Psychology.* Spetses, Greece, September 1999.

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

Núñez, M., & Harris, P. (July 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., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608–631.

Orr, H. A. (2003). Darwinian storytelling. *New York Review of Books, L*(3), 17–20.

Pereyra, L., & Nieto, J. (in press). La especificidad del razonamiento sobre situaciones de peligro. (Reasoning specializations for situations involving hazards and precautions) *Revista Mexicana de Psicología.*

Piaget, J. (1950). *The psychology of intelligence.* (M. Piercy & D. E. Berlyne, Trans.). New York: Harcourt.

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

Pinker, S. (2002). *The blank slate.* New York: Viking.

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

Price, M. E., Cosmides, L., & Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior, 23,* 203–231.

Quartz, S., & Sejnowski, T. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences, 20*(4), 537–596.

Rips, L. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Romer, P. (2000). Thinking and Feeling. *American Economic Review, 90*(2), 439–443.

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs: Implicit rules or parallel distributed processing. In D. Rumelhart, J. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 216–271). Cambridge, MA: MIT Press.

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, England: Cambridge University Press.

Shostak, M. (1981). *Nisa: The life and words of a !Kung woman.* Cambridge, MA: Harvard University Press.

Simon, H. (1990). A mechanism for social selection and successful altruism. *Science, 250,* 1665–1668.

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

Staddon, J. E. R. (1988). Learning as inference. In R. C. Bolles & M. D. Beecher (Eds.), *Evolution and learning.* Hillsdale, NJ: Erlbaum.

Stevens, J., & Stephens, D. (2004). The economic basis of cooperation: Trade-offs between selfishness and generosity. *Behavioral Ecology, 15*(2), 255–261.

Stone, V., Cosmides, L., Tooby, J., Kroll, N., & Knight, R. (2002, August). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proceedings of the National Academy of Sciences, 99*(17), 11531–11536.

Sugiyama, L., Tooby, J., & Cosmides, L. (2002, August). Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proceedings of the National Academy of Sciences, 99*(17), 11537–11542.

Tooby, J., & Cosmides, L. (1989, August). *The logic of threat.* Paper presented at the Human Behavior and Evolution Society, Evanston, IL.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 19–136). New York: Oxford University Press.

Tooby, J., & Cosmides, L. (1996). Friendship and the banker's paradox: Other pathways to the evolution of adaptations for altruism. In W. G. Runciman, J. Maynard Smith, & R. I. M. Dunbar (Eds.), *Evolution of social behaviour patterns in primates and man. Proceedings of the British Academy, 88,* 119–143.

Tooby, J., Cosmides, L., & Barrett, H. C. (2005). Resolving the debate on innate ideas: Learnability constraints and the evolved interpenetration of motivational and conceptual functions. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Structure and content.* New York: Oxford University Press.

Tooby, J., & DeVore, I. (1987). The reconstruction of hominid behavioral evolution through strategic modeling. In W. G. Kinsey (Ed.), *Primate models of hominid behavior* (pp. 183–237). New York: SUNY Press.

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

Wason, P. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth, England: Penguin.

Wason, P. (1983). Realism and rationality in the selection task. In J. St. B. T. Evans (Ed.), *Thinking and reasoning: Psychological approaches* (pp. 44–75). London: Routledge.

Wason, P., & Johnson-Laird, P. (1972). *The psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.

Wegener, J., Baare, W., Hede, A., Ramsoy, T., & Lund, T. (2004, October). *Social relative to non-social reasoning activates regions within anterior prefrontal cortex and temporal cortex.* Paper presented at Abstracts of the Society for Neuroscience, San Diego, CA.

Williams, G. (1966). *Adaptation and natural selection.* Princeton, NJ: Princeton University Press.