# The Handbook of EVOLUTIONARY PSYCHOLOGY

**Second Edition**

Volume 2: Integrations

Edited by

DAVID M. BUSS

WILEY

# The Handbook of Evolutionary Psychology

# The Handbook of Evolutionary Psychology

Second Edition

Volume 2 Integrations

Edited by
David M. Buss

WILEY

Cover design: Wiley

This book is printed on acid-free paper. ♾



Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.



***Library of Congress Cataloging-in-Publication Data:***

Handbook of evolutionary psychology (Hoboken, N.J.)
The handbook of evolutionary psychology / edited by David M. Buss. — 2nd edition.
volumes cm
Includes bibliographical references and index.
Contents: Volume 1. Foundations — volume 2. Application.
ISBN 978-1-118-75580-8 (cloth) — ISBN 978-1-118-76399-5 (set) —ISBN 978-1-118-75602-7 (pdf) — ISBN 978-1-118-75597-6 (epub)
1. Evolutionary psychology. 2. Human evolution. I. Buss, David M. II. Title.
BF698.95.H36 2016
155.7—dc23

2015008090

Printed in the United States of America

Second Edition

10 9 8 7 6 5 4 3 2 1

# Contents

# PART V

# GROUP LIVING: COOPERATION AND CONFLICT

DAVID M. BUSS and DANIEL CONROY-BEAM

HOMO SAPIENS HAS been called "the social animal" for a good reason. Living in groups defines a key mode of human existence. Groups contain a bounty of resources critical to survival and reproduction. They afford safety and protection from predators and from other humans. They are populated with potential friends for mutually beneficial social exchange. They contain reproductively valuable mates. And they are inhabited with kin, precious carriers of our genetic cargo, from whom we can receive aid and in whom we can invest. At the same time, group living intensifies competition over precisely those reproductively relevant resources, creating sources of conflict not faced by more solitary creatures. The chapters in this part describe many of the complexities of the evolutionary psychology of group living, focusing on cooperation and conflict.

In Chapter 25, Leda Cosmides and John Tooby provide a comprehensive review of the extensive body of research, much of it conducted by them and their students, on neurocognitive adaptations for social exchange. They elucidate the many design features that such adaptations theoretically should possess and provide compelling arguments that domain-general mechanisms cannot achieve the specific outcomes needed for successful social exchange. They review competing theories to explain the content effects on the Wason selection task and marshal empirical evidence relevant to adjudicating among those theories. In a display of the sort of methodological pluralism advocated by Simpson and Campbell (Chapter 3, this *Handbook*, Volume 1), Cosmides and Tooby describe cross-cultural studies, studies using traditional methods of cognitive psychology, and studies using neurocognitive techniques.

Martin Daly's chapter (Chapter 26) on interpersonal violence and homicide begins by articulating an evolutionary perspective on conflicts of reproductive interests—a long-standing ingenious strategy pioneered by Daly and his long-time collaborator Margo Wilson. Next, he articulates the rationale for using violence and homicides as assays of social conflicts. Thus, Daly's focus is not so much in explaining violence per se, although key insights into violence do indeed emerge. Rather, his central aim is to exploit patterns of violence to reveal underlying conflicts of evolutionary interests that occur between individuals when they live in groups. He deploys this strategy to make novel scientific discoveries. Kin, for example, who typically have a greater confluence of interest compared to unrelated individuals, display much less violence toward each other, despite the fact that they interact more frequently. Intimate mates, to take another example, can have converging genetic interests, as when they have mutually produced offspring. But conflicts of interest emerge from at least six sources, such as temptations for genetic cuckoldry, temptations to trade up, relationship defection, and channeling pooled resources toward one set of kin at the expense of another (see also Conroy-Beam, Goetz, & Buss, 2015). Violence is more common precisely when these conflicts of interest emerge in intimate mateships.

Anne Campbell's chapter (Chapter 27) provides an overview of theory and research on women's competition and aggression. She explores both the proximate mechanisms (hormones, physiological maturation, neuropsychology) and ultimate selective forces underlying women's competition and aggression. Fear, she argues, acts as a more powerful brake on women's than on men's violent aggression, due to the greater costs of engaging in violent conflict (e.g., costs not only to the woman, but also to her children). But make no mistake, Campbell argues—women's competition, although less ostentatiously violent, can be ferocious. Women compete for the best mates, for example, a form of competition possibly exacerbated by socially imposed monogamy. She argues that appearance (cues to fertility) and fidelity (cues to paternity certainty) become key weapons by which women compete with other women, with tactics that include shunning, stigmatizing, derogating, and ostracizing their rivals. When tactics do escalate to actual violence, they occur in predictable contexts such as resource scarcity and a sex ratio imbalance involving too few men as potential mates. In short, Campbell's excellent chapter provides a detailed analysis of the underlying adaptations for female competition and aggression, the ways in which they are sex-differentiated in design, and the contextual and ecological variables to which they respond.

Prejudice seems to be a ubiquitous feature of human social living. Everywhere, people seem prone to dislike and distrust some others, discriminating against them within groups and even warring with them when they are out-groups. Steven Neuberg and Peter DeScioli (Chapter 28) provide an outstanding chapter on the evolved psychology—threat management systems—designed to deal with adaptive problems arising from within and outside of one's group. These prejudices can cause harm and discrimination in the modern environment, they argue, which makes it all the more important to understand their design features and how they play out in this new world.

Humans are an extraordinarily coalitional species. We form groups, often in competition with other groups. Dominic Johnson's chapter (Chapter 29) on leadership and war focuses on group-on-group conflict. He outlines different hypotheses about the evolution of leader traits in the context of war, or alternatively features of coalitional leadership psychology that could have been coopted for war, and examines the relevant empirical evidence. He makes a compelling case that war has been a major selective force on human psychology, including the evolution of leadership and followership traits—arguments that have critical relevance in a modern world beset with warfare in forms unimaginable in the past, but that exploit the same suite of psychological adaptations.

Group living is what we do as a species. It offers a bounty of benefits through cooperation and an abundance of costs through social conflict. As a consequence, it is reasonable to expect that humans have evolved a large number of specialized adaptations for dealing with other humans, both for within-group interactions and for dealing with other groups. Collectively, these chapters highlight the complexity of human evolutionary psychology for group living and pave the way for the discovery of many more adaptations for grappling with the challenges posed by other humans—challenges centering on cooperation and conflict.

## REFERENCE

Conroy-Beam, D., Goetz, C., & Buss, D. M. (2015). Why do people form long-term mateships? A game-theoretic model. *Advances in Experimental Social Psychology* (Vol. 51, pp. 1–39). New York, NY: Academic Press.

CHAPTER 25

# Adaptations for Reasoning About 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; they 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. Which 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 analyses 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. There are many counterhypotheses about social exchange reasoning to test against, but they all spring from 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). According to 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. This hypothesis has been central to how most neural, psychological, and social scientists conceptualize human behavior, but 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 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 to study a form of 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, human reasoning turns 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 motivational and emotional organization found in humans (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) had sculpted the human brain and mind and, indeed, human moral reasoning. If it could be shown empirically 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). At the time, most social scientists considered morality 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, 2007; Price, Cosmides, & Tooby, 2002; Tooby & Cosmides, 2010).

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 reveal so close a fit between adaptive problem and computational solving that a neurocognitive specialization for reasoning about social exchange is implicated, including a subroutine for detecting cheaters. 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, Barrett, & Tooby, 2010; Cosmides & Tooby, 1989, 1992, 2005, 2008a; 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 transfers 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, 2007, 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, 2004; 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 counterhypothesis 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 (Dawkins, 1986; Tooby & Cosmides, 1992, Chapter 1, this *Handbook*, Volume 1; 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 (Axelrod & Hamilton, 1981; Boyd, 1988; Trivers, 1971; 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., 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

[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.

[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, Cosmides, and Tooby (2000).

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, 2008a). The following six design features (D1–D6) were among those on the list:

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, 2008a).

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.

[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.

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 conditional rules. Yet studies with the Wason selection task show that they are not. Consider the Wason task in Figure 25.1. 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 25.1: Only 5% to 30% of people give the logically correct answer, even when the rule involves familiar terms drawn from everyday life (Cosmides, 1989; Manktelow & Evans, 1979; Sugiyama et al., 2002; Wason, 1966, 1983). 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

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

Patients afflicted with Ebbinghaus disease have many different symptoms: nosebleeds, 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 25.1** The Wason Selection Task. In a Wason task, there is always a rule of the form, *If P then Q*, and four cards showing the values *P*, *not-P*, *Q*, and *not-Q* (respectively) on the side that the subject can see. From a logical point of view, only the combination of *P* and *not-Q* can violate this rule, so the correct answer is to check the *P* card (to see if it has a *not-Q* on the back), the *not-Q* card (to see if it has a *P* on the back), and no others. Few subjects answer correctly, however, when the conditional rule is descriptive (indicative), even when its content is familiar; for example, only 26% of subjects answered the above problem correctly (by choosing "has Ebbinghaus disease" and "is not forgetful"). Most choose either *P* alone, or *P* and *Q*. (The italicized *P*s and *Q*s are not in problems given to subjects.)

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 25.2a). 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 et al., 2010; Cosmides & Tooby, 1992, 1997, 2008a, 2008b; 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 disease-symptom problem in Figure 25.1. 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 25.2b; Cosmides & Tooby, 1992, 2008a; Fiddick et al., 2000). Reasoning problems could be sorted into indefinitely many categories based on their content or structure (including the propositional calculus's two content-free categories, antecedent and consequent). Yet, even in remarkably different cultures, the same mental categorization occurs. This cross-culturally recurrent dissociation by content was predicted 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 *Handbook*, Volume 1). 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 25.1 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

**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 Carters 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 25.2** 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 25.1. 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 25.4, switched social contracts).

**Table 25.1**
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, Cosmides, & Tooby, 2000). |
| B9. | That standard economic models can explain social contract effects. |
| B10. | That statistical learning produces the mechanisms that cause social contract reasoning. |

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 25.3a). 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. Familiarity is neither necessary nor sufficient for eliciting high performance (B1 of Table 25.1).

First, familiarity does not produce high levels of performance for descriptive rules (Cosmides, 1989; Manktelow & Evans, 1979). Note, for example, that the Ebbinghaus disease problem in Figure 25.1 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.

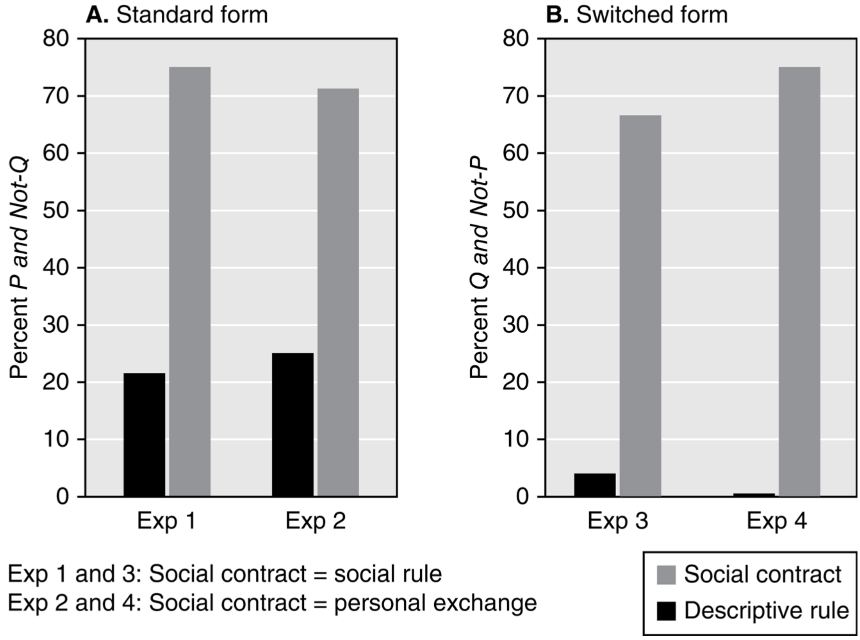


**Figure 25.3** Detecting Violations of Unfamiliar Conditional Rules: Social Contracts Versus Descriptive Rules. In these experiments, the same, unfamiliar rule was embedded either in a story that caused it to be interpreted as a social contract or in a story that caused it to be interpreted as a rule 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 25.4).

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

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 25.4** 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 (antecedents and consequents; designated here as *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 25.3B). 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.

cheaters; see Figure 25.4). 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 more than 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 requirement 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 B then C* does not imply *If C then B*; 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 more than 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 25.4). 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.*

[4] Moreover, the propositional calculus contains no rules of inference that allow *If B, then C* to be translated as *If C, then B* (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).