Adaptive specializations, social exchange, and the evolution of human intelligence

Leda Cosmides[a,1], H. Clark Barrett[a,b], and John Tooby[a]

[a]Center for Evolutionary Psychology, University of California, Santa Barbara, CA 93106; and [b]Department of Anthropology, University of California, Los Angeles, CA 90095

Blank-slate theories of human intelligence propose that reasoning is carried out by general-purpose operations applied uniformly across contents. An evolutionary approach implies a radically different model of human intelligence. The task demands of different adaptive problems select for functionally specialized problem-solving strategies, unleashing massive increases in problem-solving power for ancestrally recurrent adaptive problems. Because exchange can evolve only if cooperators can detect cheaters, we hypothesized that the human mind would be equipped with a neurocognitive system specialized for reasoning about social exchange. Whereas humans perform poorly when asked to detect violations of most conditional rules, we predicted and found a dramatic spike in performance when the rule specifies an exchange and violations correspond to cheating. According to critics, people's uncanny accuracy at detecting violations of social exchange rules does not reflect a cheater detection mechanism, but extends instead to all rules regulating when actions are permitted (deontic conditionals). Here we report experimental tests that falsify these theories by demonstrating that deontic rules as a class do not elicit the search for violations. We show that the cheater detection system functions with pinpoint accuracy, searching for violations of social exchange rules only when these are likely to reveal the presence of someone who intends to cheat. It does not search for violations of social exchange rules when these are accidental, when they do not benefit the violator, or when the situation would make cheating difficult.

evolutionary psychology | reasoning | cooperation | reciprocation

To the human mind, certain things seem intuitively correct. The world seems flat and motionless; objects seem solid rather than composed of empty space, fields, and wave functions; space seems Euclidian and 3-dimensional rather than curved and 11-dimensional. Because scientists are equipped with human minds, they often take intuitive propositions for granted and import them—unexamined—into their scientific theories. Because they seem so self-evidently true, it can take centuries before these intuitive assumptions are questioned and, under the cumulative weight of evidence, discarded in favor of counterintuitive alternatives—a spinning earth orbiting the sun, quantum mechanics, relativity.

For psychology and the cognitive sciences, the intuitive view of human intelligence and rationality—the blank-slate theory of the mind—may be just such a case of an intuition-fueled failure to grapple with evidence (1–4). According to intuition, intelligence—almost by definition—seems to be the ability to reason successfully about almost any topic. If we can reason about any content, from cabbages to kings, it seems self-evident that intelligence must operate by applying inference procedures that operate uniformly regardless of the content domains they are applied to (such procedures are *general-purpose*, *domain-general*, and *content-independent*). Consulting such intuitions, logicians and mathematicians developed content-independent formal systems over the last two centuries that operate in exactly this way. Such explicit formalization then allowed computer scientists to show how reasoning could be automatically carried out by purely "mechanical" operations (whether electronically in a computer or by cellular interactions in the brain). Accordingly, cognitive scientists began searching for cognitive programs implementing logic (5, 6), Bayes' rule (7), multiple regression (8), and other normative methods—the same content-general inferential tools that scientists themselves use for discovering what is true (7). Others proposed simpler heuristics that are more fallible than canonical rules of inference (e.g., 9, 10), but most of these were domain-general as well.

Our inferential toolbox does appear to contain a few domain-general devices (2, 11, 12), but there are strong reasons to suspect that these must be supplemented with domain-specific elements as well. Why? To begin with, much—perhaps most—human reasoning diverges wildly from what would be observed if reasoning were based on canonical formal methods. Worse, if adherence to content-independent inferential methods constituted intelligence, then equipping computers with programs implementing these methods, operating at vastly higher rates, should have made them intelligent. It did not. It turns out that general-purpose reasoning methods are very weak, and have crippling defects (e.g., combinatorial explosion) that are a direct consequence of their domain generality (1, 2, 13). Also, content effects (changes in reasoning performance based on changes in content) are ubiquitous (5, 7) yet difficult to account for in the consensus view. After all, differences in content should make little difference to procedures whose operation is designed to be content-independent. Unfortunately, the effects of content on reasoning have traditionally been dismissed as noise to be ignored rather than a window on reasoning methods of a radically different, content-specific design.

Intelligence and Evolved Specializations.

The integration of evolutionary biology with cognitive science led to a markedly different approach to investigating human intelligence and rationality: evolutionary psychology (1, 2). Organisms are engineered systems that must operate effectively in real time to solve challenging adaptive problems. The computational problems our ancestors faced were not drawn randomly from the universe of all possible problems; instead, they were densely clustered in particular, recurrent families (e.g., predator avoidance, foraging, mating) that occupy only miniscule regions of the space of possible problems. Massive efficiency gains can be achieved when different computational strategies are tailored to the task demands of different problem types. For this reason, natural selection added a diverse array of inferential specializations, each tailored to a particular, adaptively important problem domain (3). Freed from the straightjacket of a one-size-fits-all problem-solving strategy, these reasoning specializations succeed by deploying procedures that produce adaptive inferences in a specific domain, even if these operations are invalid, useless, or harmful if activated outside that

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "In the Light of Evolution IV: The Human Condition," held December 10–12, 2009, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and audio files of most presentations are available on the NAS Web site at www.nasonline.org/SACKLER_Human_Condition.

Author contributions: L.C., H.C.B., and J.T. designed research; L.C., H.C.B., and J.T. performed research; L.C. and H.C.B. analyzed data; and L.C., H.C.B., and J.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: cosmides@psych.ucsb.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0914623107/DCSupplemental.

domain. They can do this by exploiting regularities—content-specific relationships—that hold true within the problem domain, but not outside of it. This approach naturally predicts content effects, because different content domains should activate different inferential rules.

In this view, human intelligence is more powerful than machine intelligence because it contains, alongside general-purpose inferential tools, a large and diverse array of *adaptive specializations*—expert systems, equipped with proprietary problem-solving strategies that evolved to match the recurrent features of their corresponding problem domains. Indeed, the discovery of previously unknown adaptive specializations proceeded at a rapid pace once cognitive and evolutionary scientists became open to their existence and began to look for them (1).

For nearly three decades, we have been studying human reasoning in the light of evolution (14–25). By integrating results from evolutionary game theory with the ecology of hunter-gatherer life, we developed *social contract theory*: a task analysis specifying what computational properties a neurocognitive system would need to generate adaptive inferences and behavior in the social exchange problem domain (25). We have been systematically testing for the presence of these design features, including an important one: the ability to detect cheaters. Based on these investigations, we have proposed that the human mind reliably develops *social contract algorithms*: a set of programs built by natural selection for reasoning about social exchange (14–18). This system can reason adaptively about social exchange precisely because it does not perform the inferences of any standard formal logic.

Exchange as a Computational Problem

Selection Pressures for Social Exchange. Two parties can make themselves better off than they were before—thereby increasing their fitness—by exchanging things each values less for things each values more (goods, services, help). Exchange is found in every documented human culture, and takes many forms, such as returning favors, sharing food, and extending acts of help with the (implicit) expectation that they will be reciprocated. Behavioral ecology and hunter-gatherer ethnography have demonstrated that exchange is fundamental to forager subsistence and sociality (26). Indeed, evidence suggests that certain forms of social exchange were present in hominins at least two million years ago (27). This raises the possibility that selection has introduced computational elements that were well-engineered for social exchange.

Cognitive Defense Against Cheaters. Selection pressures favoring social exchange exist whenever one organism (the provisioner) can change the behavior of a target organism to the provisioner's advantage by making the target's receipt of a rationed benefit conditional on the target acting in a required manner. This contingency can be expressed as a *social contract*, a conditional rule that fits the following template: *If you accept benefit B from me, then you must satisfy my requirement R*. The social contract is offered because the provisioner expects to be better off if its conditions are satisfied [e.g., if one gives the theater owner the price of a ticket (*requirement R*) in return for access to the symphony (*benefit B*)]. The target accepts these terms only if the benefit provided more than compensates for any losses he incurs by satisfying the requirement (e.g., if hearing the symphony is worth the cost of the ticket to him). This mutual provisioning of benefits, each conditional on the other's compliance, is what is meant by social exchange or reciprocation (14, 25, 28). Understanding it requires a form of conditional reasoning that operates on abstract yet content-specific conceptual elements.

Algorithms generating social exchange require (*i*) a system for representing exchange situations in terms of conceptual primitives such as *agent*, *rationed benefit*, *provisioner's requirement*, and *entitlement*; (*ii*) a system for mapping states of the world onto these proprietary concepts (e.g., the *agent* is the theater owner; the *rationed benefit* is access to the symphony; the *requirement* is payment of the ticket price); and (*iii*) domain-specialized rules of inference that operate on these conceptual primitives, supplying inferences that are necessary to carry out the system's evolved function [rules of logic do not supply the necessary inferences (15, 17, 24)].

Adaptations for delivering benefits to unrelated individuals will be selected against unless the losses one incurs by delivering benefits are compensated for by the reciprocal delivery of benefits. Consequently, the existence of cheaters—those who fail to deliver compensatory benefits—threatens the evolution of exchange. Using evolutionary game theory, it has been shown that adaptations for social exchange can be favored and stably maintained by natural selection, but *only if they include design features that enable them to detect cheaters*, and cause them to channel future exchanges to reciprocators and away from cheaters (28–30).

Cheater Detection Is Person Categorization and Not Event Categorization. Indeed, the exact nature of the adaptive problem posed by cheaters determines how social contract algorithms should have evolved to define the concept *cheater*. A cheater is an agent who (*i*) takes the rationed benefit offered in a social exchange but (*ii*) fails to meet the provisioner's requirement, and (*iii*) does so by intention rather than by mistake or accident. The evolutionary function of a cheater detection subroutine is to defend the cooperator against exploitation. It is designed to represent and track the other party's behavior so that it can (when warranted) correctly connect an attributed disposition (to cheat) with that particular person (who thereby becomes categorized as a *cheater*). In the experiments reported here, we manipulate these cheater-defining elements and demonstrate that they regulate reasoning about conditional rules involving social exchange.

The conceptual primitive *cheater*, along with domain-specific inference rules for detecting cheaters, were predicted to be a central part of social exchange reasoning because they are necessary to solve a computational problem that, if unsolved, would prevent the evolution of social exchange. Together, they constitute a logic of social exchange—a content-specialized logic whose conceptual primitives, inference rules, and outputs are quite different from those produced by standard formal logics, such as the predicate calculus. The need for special inferential rules for cheater detection derives from the fact that standard, domain-general conditional reasoning rules will fail to identify cheaters in many circumstances, and will misidentify reciprocators and altruists as cheaters in others (15–18).

Evolutionary Versus Economic and Other Functions. A cheater is someone who has violated a social contract—a conditional rule involving social exchange—but not all violations of social contracts reveal the presence of a cheater. This fact allows critical tests between an evolutionary and an economic perspective.

A commonplace of economics, utility theory, and common sense is that people become adept at solving problems that they are motivated to solve, such as those that have economic consequences. When targets violate a social contract, the provisioner suffers a loss in utility; if these cases are detected, the provisioner could insist on getting what she is owed. If such economic consequences drive the acquisition of reasoning skills, then people should be good at detecting all violations of social contracts because this will allow them to recoup their losses. It would not matter whether the violation occurred by mistake or on purpose, or whether it benefited the violator—the provisioner has suffered a loss in all of these cases. These distinctions do matter, however, if evolution produced a subroutine specialized for detecting cheaters.

The function of a cheater detection subroutine is social categorization: this person = cheater. The fitness benefit that drove the evolution of cheater detection is the ability to avoid squandering costly future cooperative efforts on those who will exploit rather than reciprocate. Violations of social contracts are relevant to this

evolved function, but only insofar as they reveal individuals disposed to cheat—that is, individuals who cheat by virtue of their calibration or design. Noncompliance caused by accidental violations and other innocent mistakes do not reveal the disposition to cheat. Hence, they should not be encoded by social contract algorithms as cheating—even though the payoff to the provisioner is the same as cheating. That is, accidental violations may result in someone being cheated (not getting what they are entitled to), but they do not indicate the ongoing menace of a cheater.

Therefore, social contract theory predicts cognitive design features beyond detecting compliance or noncompliance with a social contract. The subroutine should be designed to look for potentially intentional violations, because only these predict future defection and continued exploitation—the negative outcome the system evolved to defend itself against. Indeed, results from evolutionary game theory show that strategies for conditional cooperation do better if they can "see through" failures to reciprocate due to chance, because they continue to reap the benefits of cooperating with other conditional cooperators who may have erred (e.g., 31). The surprising social contract prediction—that noncompliance in social exchanges is only detected at high rates when it could reveal cheaters—is tested in the experiments reported below.

Cue-Based Activation of Cheater Detection. To achieve their large efficiency gains, adaptive specializations should be designed to be differentially activated when they encounter the content domains they are equipped to solve, and inactive when they encounter other content domains where their proprietary operations are inapplicable, mismatched, or invalid (1, 13). This point is central to understanding the logic of the experiments reported here: If there is an evolved inference system specialized for reasoning about social exchange, then the cheater detection subroutine should be differentially activated by content cues signaling the potential for determining whether someone is a cheater. The strategy pursued in these experiments is to add and subtract minimal problem elements that are irrelevant to competing theories of reasoning but that should activate or inactivate the cheater detection subroutine because they allow or interfere with the determination of whether someone is a cheater. The elements we will concentrate on are: (*i*) Is there a *benefit* being rationed? (if not, there is no social exchange); (*ii*) Could the other party benefit by violating of the rule? (if not, then detecting a violation will not identify a cheater); (*iii*) Did the violator have the intention to cheat? (if not, then detecting a violation will not identify a cheater); and (*iv*) Does the situation make cheating difficult? (if so, then looking for violations is unlikely to reveal cheaters).

Experimental Tests of Social Contract Theory

Exchange is, by definition, social behavior that is conditional: The provisioner agrees to deliver a benefit *conditionally* (conditional on the recipient doing what the provisioner required). Engaging in social exchange therefore requires conditional reasoning.

Our experiments use the Wason selection task, a standard tool for investigating conditional reasoning (5). Subjects are given a conditional rule of the form *if P then Q*, and asked to identify possible violations of it—a format that allows one to see how performance varies as a function of the rule's content (Fig. 1). It was originally developed to determine whether humans are natural falsificationists—whether the brain spontaneously applies first-order logic to look for cases that might violate a hypothesis or other conditional rule. To their surprise, psychologists found that people perform poorly when asked to solve this simple problem.

According to first-order logic, a conditional rule has the form *if antecedent* (conventionally represented by *P*) *then consequent* (conventionally represented by *Q*). Looking for violations of a conditional rule is a remarkably simple task: According to first-order logic, the rule is violated whenever *P* is true but *Q* is not—that is, by the co-occurrence of *P & not-Q*. For example, the rule "if a person is a biologist, then that person enjoys camping" would be

The following rule holds: *"If P then Q"*. The cards below have information about four situations. Each card represents one situation. One side of a card tells whether *P* happened, and the other side of the card tells whether *Q* happened. Indicate only those card(s) you definitely need to turn over to see if any of these situations violate the rule.

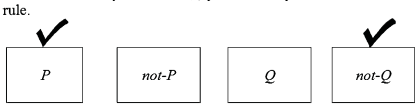

Fig. 1. The general structure of a Wason selection task. The rule always has specific content; e.g., "if a person is a biologist, then that person enjoys camping" (an indicative rule). For this rule, each card would represent a different person, reading (e.g.) "biologist" (*P*), "chemist" (*not-P*), "enjoys camping" (*Q*), "does not enjoy camping" (*not-Q*). The content of the rule can be varied such that the rule is indicative, a social contract, a precaution, a permission rule, or any other conditional of interest, allowing alternative theories of reasoning to be tested. Check marks indicate the logically correct card choices.

violated by finding a biologist who does not enjoy camping. Thus, a fully correct Wason response would be to choose *P*, *not-Q*, and no other cards. First-order logic is simple to specify—vastly simpler than many other cognitive capacities that humans are known to have, such as vision and grammar acquisition. It is also prototypically content-independent and domain-general: It maps all of the content of conditional rules into the format *if antecedent then consequent*, where the antecedent and consequent can stand for any propositions. It can be informative about nearly everything. Finally, the inferential rules of first-order logic are incapable of responding differentially to particular contents, because they do not represent content at all. All its rules see are its conceptual primitives (e.g., *antecedent*, *consequent*, *if*, *then*).

The biologist-camping rule is *indicative*: It claims to indicate or describe some relationship in the world. Given an indicative conditional, only 5–30% of normal subjects respond with the logically correct answer, *P & not-Q*. Although one might think that people would learn to reason correctly about familiar relationships, performance remains poor even when the indicative rule involves very familiar content drawn from direct experience—such as, *if I eat cereal, then I drink orange juice* (14, 32). Either our species did not evolve a full and unimpaired version of first-order logic, or significant parts of it are not activated when people try to solve this simple information search task.

Although originally designed to test whether people had a content-general system for conditional reasoning, the Wason task can be and has been used to test many of the predictions of social contract theory (for reviews, see refs. 16 and 17). The hypothesis that the brain contains social contract algorithms predicts that reasoning performance should shift when it encounters social exchange content. In particular, it predicts a sharply enhanced ability to reason adaptively about conditional rules when those rules specify a social exchange.

Content is indeed decisive. Although performance on the Wason selection task is typically poor, when the conditional rule involves social exchange and detecting a violation corresponds to looking for cheaters, 65–80% of subjects correctly detect violations (14–18). Confounding the idea that people improve only with increasing experience, subjects perform well even when the conditional rule specifies a wildly unfamiliar social contract that no subject has ever heard before (e.g., "if you get a tattoo on your face, then I'll give you cassava root"). Indeed, we have found no difference in performance between totally unfamiliar social contracts and thoroughly familiar ones. Moreover, the ability to detect cheaters on social contracts is present as early as it can be tested (age 3–4) (33). Consistent with the hypothesis that this adaptive specialization is part of our species' cognitive architecture, this pattern of results has been found in every culture where it has been tested, from industrialized market economies to Shiwiar hunter-horticulturalists of the Ecuadorian Amazon (23).

Because the correct answer if one is looking for cheaters is sometimes the logically correct answer as well, many people mis-

understand us to be claiming that social exchange content boosts *logical* reasoning. But social exchange activates nonstandard rules of inference that diverge sharply from first-order logic. The adaptively correct answer when one is looking for cheaters is to choose cards representing people who have *taken the benefit* and *not met the requirement*, regardless of the logical category into which these fall. It is possible to create social exchange problems in which these correspond to a logically incorrect answer: *Q & not-P*. When this was done, subjects' selections matched the predictions of the nonstandard evolutionary logic of social exchange and violated first-order logic. *Q* and *not-P* choices were predicted from first principles, and had never been predicted by any other theory, or previously elicited (14, 15, 18).

Deontic Logic or Social Contract Algorithms? Most reasoning researchers now concede that (*i*) content effects for social exchanges are real, replicable, and robust, and (*ii*) first-order logic cannot explain how people reason about social exchange. However, they continue to resist the counterintuitive notion that reasoning in this domain is governed by an adaptive specialization. Instead, they propose that reasoning about social exchange is governed by some version of deontic logic—a formal system for reasoning about concepts such as permission and obligation. Social contracts do involve deontic concepts, so this is a plausible proposal. However, the universe of deontic rules is far larger than social contracts, extending to moral rules, imperatives, norms, and so on. If the mind comes equipped with (or acquires) some form of deontic logic, then people should be good at detecting violations of all deontic rules (34, 35), or at least those involving utilities (36–38).

In support of the deontic position, its advocates point out that another type of deontic conditional that we and others have worked on also elicits good violation detection on the Wason task: precautionary rules (19, 20, 24, 36, 39). These are conditionals that fit the template "*if one is to engage in hazardous activity H, then one must take precaution P*" (e.g., "if you are working with toxic gases, then you must wear a gas mask"). Precautionary rules are so similar to social contracts that most theories view them as trivial variations on a theme—deontic conditionals involving utilities, processed by precisely the same reasoning system (36–41).

By contrast, evolutionary researchers have proposed that these rules are processed not by social contract algorithms but by a different specialization with a distinct function: to monitor for cases in which people are in danger by virtue of not having taken the appropriate precaution (24, 42). Neuroimaging results and evidence that brain damage can selectively impair social exchange reasoning (while sparing precautionary reasoning) support the evolutionary hypothesis that these are two distinct specializations, not one superordinate deontic system (19–22). Moreover, an adaptationist perspective predicts that subjects should detect accidental as well as intentional violations of precautionary rules, because both place people in danger—a prediction now tested and confirmed (43). By contrast, intentionality should matter in exchanges, because unintentional violations do not reveal cheaters tnqh_x2014;a prediction we test below.

New Tests Between Deontic and Social Exchange Theories. In exchange, an agent *permits* another party to take a benefit, conditional upon that party's having met the agent's requirement. There are, however, many situations other than exchange in which an action is permitted conditionally. A *permission rule* is any deontic conditional that fits the template "if one is to take action A, then one must satisfy precondition R" (34, 39). According to Cheng and Holyoak's permission schema theory (34), reasoning about such rules is governed by four production rules that result in people checking the "action taken" card (*P*) and the "precondition not met" card (*not-Q*). We focus on their account because it is the most precisely specified of the deontic accounts. But our experiments also test against all other deontic theories known to us.

It is important to remember that all social contracts are permission rules, but there are many permission rules that are not social contracts. According to permission schema theory, good violation detection is elicited by the entire class of permission rules (34, 39)—a far more inclusive and general set that includes precautionary rules, bureaucratic rules, etiquette rules, and social norms, along with social contracts.

Empirically, permission schema theory is undermined or falsified if permission rules that are neither social contracts nor precautions routinely fail to elicit high levels of violation detection. Moreover, social contract theory is supported if social exchange rules (a subset of permission rules) fail to elicit this effect when violation detection does not identify cheaters.

Just how precise and functionally specialized is the reasoning system that causes cheater detection? Below we report the results of four experiments that parametrically test predictions about benefits, intentionality, and the ability to cheat on social contracts. The full text of all Wason tasks tested is given in *SI Text*. All experiments had a between-subjects design in which each subject was given a single Wason task. In every case, the rule tested was a *permission rule*: a deontic conditional that fits the template "if one is to take action A, then one must satisfy precondition R." For the problems tested, the correct answer if one is looking for violations of the rule is to choose the *P* card (action taken), the *not-Q* card (requirement not met), and no others (henceforth *P & not-Q*).

Permission Rules Without Benefits

The function of a social exchange for each participant is to gain access to a benefit that would otherwise be unavailable to them. Therefore, an important cue that a conditional rule is a social contract is the presence of a desired benefit under the control of an agent—this should activate social contract reasoning. Experiments 1 and 2 compare performance on permission rules that vary a single element: whether *P*, the action to be taken, is a benefit to the potential violator. They show that this has a dramatic effect on performance.

Experiment 1. Wason tasks consist of a rule and a story setting the context (what the cards refer to, who proposed the rule, and so on). In experiment 1 we kept the stories identical, and made small but theoretically relevant alterations to the rule. By changing *P* to an action that our subjects would spontaneously recognize as a benefit, a permission rule was transformed into a social contract. A two-word change was sufficient (see below, rules 1 and 2).

Experiment 1 compared performance on three permission rules, each described as a law made by a group of tribal elders (*SI Text*). The story context, which was minimal, was identical for rules 1 and 2, and nearly so for rule 3. The four cards represented what four different members of the tribe had done, and subjects were asked which card(s) they would need to turn over to see whether any of these people had broken the law.

The tribal context allowed the use of permission rules that would be unfamiliar to the subjects (no underlining in the originals):

1. "If one is going out at night, then one must tie a small piece of red volcanic rock around one's ankle."
2. "If one is staying home at night, then one must tie a small piece of red volcanic rock around one's ankle."
3. "If one is taking out the garbage, then one must tie a small piece of red volcanic rock around one's ankle."

The underlined portion of these permission rules specifies an action that one is permitted to take only if a requirement is met (the red rock is worn). What varies is whether this action is a benefit to the potential violator. Our subjects were undergraduates, for whom going out at night (rule 1) is a highly favored activity. Staying home at night (rule 2) is usually neutral: it is what

you do when you have too much work, do not have a date, etc. Taking out garbage (rule 3) is a mildly unpleasant chore.

From the standpoint of permission schema theory, there is no theoretical difference between rules 1–3. Permission rules regulate when an action can be taken—whether that action benefits the potential violator or not. Cheng and Holyoak (39) are very clear on this point; indeed, it is the basis for their claim that high performance on precautionary rules (where the action taken is hazardous and unpleasant) supports the existence of a permission schema. If permission schema theory is correct, all three rules will elicit high levels of violation detection, resulting in equally high percentages of *P & not-Q* responses.

Benefits play no role in permission schema theory, but they play a key role in social contract theory. Because rule 1 regulates access to an activity that our subjects recognize as a benefit, it should be interpreted as a social contract. This should activate the cheater detection mechanism, leading to a high level of *P & not-Q* responses. In contrast, rules 2 and 3 do not regulate access to a benefit. Lacking this key element, they are less likely to be interpreted as social contracts, and less likely to activate cheater detection. For this reason, social contract theory predicts that rules 2 and 3 will yield a significantly lower percentage of *P & not-Q* responses than rule 1.

That is precisely what happened. Rule 1—the social contract—elicited the correct response (*P & not-Q*) from 80% of subjects (20/25). Only 52% of subjects (13/25) answered correctly for rule 2, and 44% (11/25) for rule 3. Both of these percentages are significantly lower than that found for the social contract (80% vs. 52%: $Z = 2.09$, $P = 0.018$, $\phi = 0.30$; 80% vs. 44%: $Z = 2.62$, $P = 0.0044$, $\phi = 0.37$). Rules 1 and 2 are true minimal pairs: They differ by only two words (going out vs. staying home)—a trivial difference of no importance for permission schema theory. Yet this two-word difference, which affects whether the antecedent is a benefit or not, caused a drop in performance of 26 percentage points. The garbage problem was included because the less pleasant the activity being regulated, the lower the probability that subjects will interpret the rule as a social contract; this predicts a monotonic decrease in performance from going out to staying home to garbage. Contrast analysis confirms this (several monotonically decreasing models work, with $\lambda = +3, -1, -2$ providing the best fit, $Z = 3.06$, $P = 0.0011$; *SI Text*, note 1).

Note that rule 3, the garbage rule, is *capable* of eliciting high levels of violation detection. In a parallel study, 72% of subjects (18/25) answered correctly when the garbage rule was embedded in a story indicating that the people of this tribe view it as precautionary—that is, when taking out the garbage involves hazards that can be avoided by wearing the rock [72% (precautionary) vs. 44% (not precautionary): $Z = 2.01$, $P = 0.02$, $\phi = 0.28$; *SI Text*, experiment 5]. This illustrates an important point: High levels of violation detection are typically found for social contracts and for precautionary rules, but not for permission rules that fall outside these categories. The garbage and staying home problems in experiment 1 elicited poorer performance because they were not social contracts *or* precautionary.

What about the requirement term? A rule is a social contract if it regulates access to a benefit; it does not matter whether the requirement the individual must satisfy to be entitled to that benefit is personally costly, neutral, or even beneficial. The requirement is imposed not because it is costly to the person who must satisfy it, but because doing so creates a situation that *the agent who imposed it* wants to achieve. In a separate study, we confirmed this prediction: Varying whether the requirement is costly or beneficial to the potential violator had no effect on performance (*SI Text*, experiment 1-A).

Experiment 2. Experiment 2 is a conceptual replication of experiment 1, using the 1970s-era Sears problem.

4. "You work as an assistant at Sears. You have the job of checking sales receipts to make sure that any sale over $30 has been approved by the section manager (this is a rule of the store)."

The cards, which have information about four sales receipts, read "$70," "$15," "signed," and "not signed" (*SI Text*). D'Andrade (in ref. 44) reports performance of ~70% on this problem.

According to Cosmides (14, 15), performance on the Sears problem is high because it engages the social contract algorithms. Sears is an institution devoted to social exchange, with policies to prevent cheating—checking credit cards and driver's licenses, asking for phone numbers on checks, and so on. The more expensive the item, the worse it is for the store when people do not pay. Requiring manager approval for expensive items is one way stores protect themselves from potential cheaters: The manager approves the expensive purchase only when there are indications that the customer will be able to pay. The rule therefore regulates access to a benefit: goods worth more than $30 versus those worth less. Others dispute this interpretation, arguing that the Sears problem is a permission rule but not a social contract, and that high performance on this problem demonstrates the existence of a reasoning mechanism general to the class of deontic rules (36, 39).

There is no need to engage in verbal arguments when these two views can be tested empirically by comparing performance on rule 4, the original Sears problem, with closely matching rules that do not regulate access to a benefit. To this end, we created permission rules that are about inventory forms rather than sales receipts, such as:

5. "You work as an assistant at Sears. Each department at Sears (Menswear, Sportswear, Ladies Shoes, etc.) has a different color inventory form to fill out. You have the job of checking inventory forms that the department clerks have filled out to make sure that any blue inventory form has been signed by the section manager. (This is a rule of the store.)" (Cards read: "blue," "white," "signed," "not signed.")

By making the rule about inventory forms, we have created a permission rule that subjects are unlikely to interpret as a rule regulating access to a benefit: It is unclear how anyone might benefit by filling out a blue inventory form rather than a white one. If high performance on rule 4 was caused by social contract algorithms, then performance on rule 5 should be lower. By contrast, a permission schema should generate equally high performance on rules 4 and 5.

Contrary to the predictions of permission schema theory, performance on the Sears inventory problem was significantly lower than performance on the original Sears problem: 72% correct (18/25) on rule 4 versus 48% correct (12/25) on rule 5 ($Z = 1.73$, $P = 0.04$, $\phi = 0.24$). Can it be pushed even lower by removing elements relevant to social exchange?

Most people know that requiring signatures is a common device to protect against cheating; the fact that a manager's signature is required for blue inventory forms could suggest to some that valuable goods are being tracked by the blue forms, but not the white ones. To remove any hint that *P* might represent access to something more valuable than *not-P*, we created rule 6, an inventory rule that regulates where blue forms are to be filed (*SI Text*).

Performance was even lower in response to rule 6: 32% correct (8/25). Yet it is a permission rule set in a culturally familiar context. Consistent with the view that performance should decrease with the removal of cues suggesting a benefit is being regulated, contrast analysis confirms a monotonic decrease from sales receipts (72%) to signed inventory forms (48%) to filed inventory forms (32%) (with $P \leq 0.0012$, $Z \geq 3.04$ for $\lambda = 3, -1, -2$ and $1, 0, -1$).

These results support social contract theory and disconfirm permission schema theory. Indeed, they refute any theory [including Fodor's (35)] that attributes good violation detection to the entire domain of deontic rules.

Cheater detection has a signature: *Benefit accepted* and *requirement not met* are chosen, regardless of logical category. That the conditional rule regulates access to a benefit is a necessary condition for eliciting the detection of this very precise type of violation (15, 24). It is not, however, a sufficient condition (18)—a prediction that we test in experiments 3 and 4.

Intentional Violations Versus Innocent Mistakes

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. Other deontic theories make the same prediction, because their explanations rest on how the rule is interpreted, not on properties of the violator.

By contrast, the evolved function of a cheater detection subroutine is to use cues of an intentional failure to meet the requirement to correctly connect an attributed disposition (to cheat) with a person (a cheater), for the reasons outlined above. Accidental violations may result in someone not getting what they are entitled to, but without indicating the presence of a cheater. Social contract theory therefore predicts that the same social contract rule will elicit lower violation detection when the context suggests that violations were occurring by mistake rather than by intention (18, 25). One partial test showed lowered violation detection when the individuals who might mistakenly violate the social contract were not in a position to obtain the benefit regulated by it (43). However, as the benefits experiments above make clear, this lower performance might have been due to the lack of a benefit to the violator, and not to intentionality at all.

We designed experiments 3 and 4 to clearly test whether intentionality regulates violation detection in social exchanges and, if so, to pinpoint exactly what cues activate or inactivate social contract algorithms. If the cheater detection mechanism is designed to check whether the potential violator has obtained the benefit specified in the rule, then we would see a drop in performance when the violator will not obtain that benefit—even if his violation was intentional. Another possibility is that the cheater detection mechanism is not activated by innocent mistakes—even when the rule violator gets the benefit regulated by the rule by making this mistake. A third possibility is that the social exchange system is designed to respond to both of these factors.

Experiment 3: Intentionality Without Benefits. In this study, the social contract rule was held constant; what varied was whether the potential violators were (*i*) cheaters (individuals who intend to violate the rule to obtain the regulated benefit), (*ii*) saboteurs (individuals who intend to violate the rule, but do so to obtain another benefit, rather than the benefit regulated by the rule), or (*iii*) people who may have made innocent mistakes (no intention to violate, no benefit gained by doing so).

The social contract regulated access to a very high quality school, Dover High. The story explains that it is a great school with an excellent record for placing students in good colleges; the neighboring school, Hanover High, is mediocre, with poor teachers and decrepit facilities. The story further explains that Dover High is good because the people of Dover City pay high taxes to support it; in contrast, the equally prosperous people of Hanover and Belmont (neighboring communities) have not been willing to spend the money it would take to improve Hanover High.

Taking these factors into account, the Board of Education made rule 7:

7. "If a student is to be assigned to Dover High School, then that student must live in Dover City."

In all versions, subjects are asked to imagine that they supervise four volunteers at the Board of Education who are supposed to follow rule 7. Each card represents the documents of one student, and the subject is asked which they need to turn over to see whether the documents of any of these students violate the rule.

In the cheater condition, each volunteer is the mother of a teenager who is about to enter high school, and each processed her own child's documents; the concern is that some might have cheated. In the innocent mistake condition, the volunteers are helpful elderly ladies who have become absent-minded; the concern is that some might have violated the rule by mistake. In the sabotage condition, the volunteers are mad at you for having fired their best friend; the concern is that they intend to violate the rule, with the goal of creating chaos that will make you look incompetent in the eyes of your boss.

Performance was best in the cheater condition and worst in the innocent mistake condition: 68% versus 27% correct (23/34 vs. 9/33; $P = 0.0005$, $\phi = 0.40$). The sabotage condition elicited intermediate performance of 45% correct (15/33)—significantly worse than the cheater condition ($P = 0.033$, $\phi = 0.22$) and marginally better than the innocent mistake condition ($P = 0.06$, $\phi = 0.19$). Contrast analysis confirms a linear decrease in performance in the cheater, sabotage, and innocent mistake conditions ($\lambda = +1, 0, -1$: $Z = 3.62$, $P = 0.00015$), and performance on all three was significantly better than on two other permission rules tested that were not social contracts at all, one using rule 7 ($10^{-6} < P\text{s} < 0.047$; *SI Text*).

The results indicate that the cheater detection mechanism is most strongly activated by situations suggesting there are individuals who (*i*) *intend* to violate a social contract rule, and (*ii*) will *gain the benefit* it regulates by doing so. Removing the ability to gain this benefit while retaining the intention to violate decreased performance by ~20 percentage points; removing both factors decreased performance by ~40 percentage points.

Experiment 4 tests this interpretation in a full parametric study. It tests whether the effects of these variables are independent, and whether accidental violations decrease performance even when the violator *will* benefit from her mistake.

Experiment 4: Manipulating Intentions, Benefits, and Ability. Eight Wason problems were tested, all using the same social contract—a version of rule 7 ("If a student is to be assigned to Grover High School, then that student must live in Grover City"). The story provided the same explanation as before: Grover High is better than Hanover High, and access to this benefit is restricted to students living in the community that pays higher taxes to support it (*SI Text*).

Everything about the problems was held as constant as possible, while three variables were manipulated: (*i*) whether the volunteers intended to violate the rule (*intention* present vs. absent); (*ii*) whether the volunteers could gain the benefit regulated by the rule by violating it (*benefit* present vs. absent), and (*iii*) whether the *ability* to easily cheat was present or absent. These three variables were fully crossed in a $2 \times 2 \times 2$, between-subjects design.

The search for rule violations is unlikely to reveal individuals with a disposition to cheat when the situation prevents cheating. The *ability* variable was included to see whether cheater detection is relaxed under these circumstances. In real life, some situations preclude cheating (18), either by both parties (e.g., simultaneous exchange of goods; working jointly on a shared project) or by one party (e.g., having done a favor in advance of reciprocation precludes cheating by the favor doer). In other situations, measures are taken to make cheating more difficult. To simulate the latter situation, the *ability-absent* conditions said that students were identified by code numbers so that the volunteers could not know which documents belonged to which child. This information was omitted from the *ability-present* conditions.

As above, *benefit-present* conditions explained that each volunteer is the mother of a teenager who is about to enter high school, and each processed her own child's documents. *Benefit-absent* conditions said none of the four volunteers have children in school

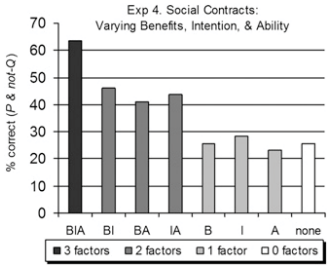

Fig. 2. Parametric study of social contract reasoning. In all conditions, subjects were asked to look for violations of the same social contract. What varied was whether potential violators could *benefit* by violating the rule (B), whether their violating the rule was by *intention* or by mistake (I), and whether the situation provided them with the *ability* to easily violate it (A). When all three factors were present (BIA), performance was highest. It dropped significantly when only two factors were present (BI, BA, IA), and again when only one factor was present (B, I, A), to the same low level found when none of these factors were present.

and so could not benefit from misassigning students. *Intention-present* conditions said you (the supervisor) overheard that some of your volunteers intended to try to break the rules when it came to assigning children to schools (or "mischievously intended," to create a motive for this intention in the *benefit-absent* conditions). *Intention-absent* conditions explained that you believe your volunteers are honest, but suspect they may have made some innocent mistakes and broken the rules for assigning each child to a particular school. The percentage of subjects correctly detecting violations by choosing the Grover High card (*P*), the town of Hanover card (*not-Q*), and no others is shown for each condition in Fig. 2.

The results were remarkably clear: Each factor—benefit, intention, and ability—contributed to violation detection, independently and additively (three-way ANOVA, main effects: *Benefit* $F_{1,342} = 7.30$, $P = 0.007$, $\eta = 0.14$; *Intention* $F_{1,342} = 10.22$, $P = 0.002$, $\eta = 0.17$; *Ability* $F_{1,342} = 4.87$, $P = 0.028$, $\eta = 0.12$; no interactions). The percentage of subjects who answered correctly was 64% when all three factors were present (BIA), 46% when only two factors were present (BI, BA, or IA), and 26% when only one factor was present (B, I, or A)—the same performance found when no factors were present. That is, each time a factor was removed, performance dropped by about 20 percentage points. This is the same pattern found in experiment 3, where *ability* to cheat was always present: 68% correct for the three-factor cheater condition (BIA), 45% correct for the two-factor sabotage condition (IA), and 27% correct for the one-factor innocent mistake condition (A) (*SI Text*, note 2).

Discussion

Problems with Deontic and Other Theories. Experiments 1–4 falsify every deontic theory that we are aware of: ones that apply to the entire class of deontic rules, as well as ones that apply only to those involving utilities. Why? If high performance on social contracts were due to the ability to do well on deontic rules, then subjects should routinely do well on problems where the deontic rules are neither social contracts nor precautions. They do not. In all cases tested in these experiments, subjects did poorly on deontic rules that did not allow the detection of a cheater. This was true even for social contracts, which are deontic permission rules involving utilities: Performance was high when detecting violations would reveal cheaters, but low when it would not.

Interpretation Theories Ruled Out. Most theories attempting to explain the spike in violation detection for deontic rules focus on how the rule itself is interpreted. These theories claim that violations are detected because deontic conditionals are interpreted as implying either (*i*) the inferences of a permission schema (34); (*ii*) the slightly more general (but otherwise identical) inferences of the material conditional in first-order logic (45); (*iii*) that cases of *P & not-Q* are forbidden [relevance theory (38)]; or (*iv*) that *Q* is required (35). None of these theories can explain the results of experiments 1–4 because, although every rule tested was clearly framed as a deontic conditional, most failed to elicit good violation detection.

This is most strikingly demonstrated by the results of experiment 4, where the same social contract, with the same interpretation, was used in every condition. Despite holding the rule and its interpretation constant, we predicted and found systematic decreases in violation detection. This was accomplished by subtracting cues that the search for violations might reveal people with a disposition to cheat. Each cue removed dropped performance by ~20 percentage points, from a high of 64% (three cues) to 46% (two cues) to 26% (one cue).

Economic and Utility Consequences Ruled Out. Economics, utility theory, behaviorism, and common sense all lead to the expectation that a lifetime of not getting what you are entitled to would build skill at detecting violations of social contracts. Yet our subjects were not good at detecting these violations unless doing so might reveal a cheater—someone who had intentionally taken the benefit without meeting the requirement.

Some reasoning researchers take a similar tack. To explain why social contracts and precautionary rules—but not other deontic conditionals—elicit good violation detection, they propose that the deontic rule must involve utilities, and that a rule violation must have consequences for someone's utility (e.g., ref. 37; refs. 40 and 41 on optimal data selection and decision theory; ref. 38 on relevance theory). However, the results of experiments 3 and 4 cannot be explained by these theories either. The school problem was a social contract and, whether it is violated by accident or intent, in all cases violations will produce an event with consequences for other people's utility: Someone's child will get access to a benefit they are not entitled to, and the school board (and taxpayers) will experience a loss—they will incur the expense of providing a benefit to families that did not pay for it (ref. 46, and *SI Text*, note 3).

The most dramatic demonstration that "consequences for utility" is an inadequate explanation comes from comparing the four *benefit-present* conditions of experiment 4 (BIA, BI, BA, B; Fig. 2). A rule violation has obvious consequences for utility in all of these cases: A volunteer will benefit if her child is assigned to a good school that she did not pay for—a point which is brought to the subject's attention (*SI Text*, note 2). Yet performance decreased as a function of *intention* and *ability*, from 64% in the BIA condition, to 41% and 46% in the BA (*intention* removed) and BI (*ability* removed) conditions, to 26% in the B condition (*intention* and *ability* removed). Indeed, the relevance [sensu Sperber et al. (38)] of discovering violations is highest in the BI condition, because this would reveal that the anticheating measures taken are ineffective, that people are intentionally breaking the rule despite these measures, and that they are getting an unearned benefit by doing so. Yet performance in this condition was *lower* than for the BIA condition, and similar to performance in the BA condition, where there was no intention to cheat.

Lastly, the deontic + utility theories cannot explain the results by claiming—post hoc—that violation detection procedures are simply not activated by the prospect of accidents and other innocent mistakes. If this were true, then violation detection should suffer when people are asked to look for accidental violations of precautionary rules; like social contracts, these are deontic conditionals involving utilities. Yet the accident-intention manipulation has no effect whatsoever on precautionary rules: People easily detect accidental violations of them (43). This is what one would expect if they were being processed by the precautionary system—an adaptive specialization with a different evolved function.

The Remarkable Functional Specificity of Cheater Detection Algorithms. Progress is made not only by ruling out rival theories but by mapping additional design features in the architecture of social contract algorithms. These experiments clarify a number of design features of the social contract inferential specialization. Based on the distinctive pattern in which violation detection is up-regulated and down-regulated, we now know that at least three cues independently contribute to cheater detection. First, intentional violations activate cheater detection, but innocent mistakes do not. Second, violation detection is up-regulated when potential violators would get the benefit regulated by the rule, and down-regulated when they would not. Third, cheater detection is down-regulated when the situation makes cheating difficult—when violations are unlikely, the search for them is unlikely to reveal those with a disposition to cheat. This provides three converging lines of evidence that the mechanism implicated is not designed to look for general rule violators, or deontic rule violators, or violators of social contracts, or even cases in which someone has been cheated; it does not deign to look for violators of social exchange rules in cases of mistake—not even in cases when someone has accidentally benefited by violating a social contract. Instead, this Inspector Javert-like system is monomaniacally focused on looking for social contract rule violations when this is likely to lead to detecting cheaters—defined as agents who obtain a rationed benefit while intentionally not meeting the requirement.

Consider, however, that this system has the computational power to detect social contract violations—it must, to detect cheaters. Yet it is regulated so that violation detection is deployed only in the service of cheater detection. It is difficult to think of a more powerful signal of the system's functional specificity than that.

Conclusions

An evolutionary approach to human intelligence leads to the radically different—and highly counterintuitive—view that our cognitive architecture includes evolved reasoning programs that were specialized by selection for distinct adaptive problems, such as social exchange and evading hazards. Although such a view strikes many as implausible in the extreme—why would anything in the mind be so strangely specialized?—the careful analysis of adaptive problems allows the derivation of rich sets of testable and unique predictions about our cognitive architecture. When these predictions are empirically tested—as here—the results typically support the view that the human cognitive architecture contains specializations for adaptive problems our ancestors faced. In this case, we can show that the cheater detection system functions with pinpoint accuracy, remaining inactive not only on rules outside the domain of social exchange but also on social exchanges that show little promise of revealing a cheater. In the contest of intuition versus evidence, it will be interesting to see which will prove the more persuasive. Human intelligence, like the sediments of east Africa, may preserve powerful signals from the evolutionary past. And, maybe, the Earth really does orbit the sun.

ACKNOWLEDGMENTS. The authors thank Roger N. Shepard and the NIH Directors' Pioneer Award to L.C.

1. Tooby J, Cosmides L (1992) *The Adapted Mind*, eds Barkow J, Cosmides L, Tooby J (Oxford Univ Press, New York), pp 19–136.
2. Cosmides L, Tooby J (2001) *The Evolution of Intelligence*, eds Sternberg R, Kaufman J (Erlbaum, Hillsdale, NJ), pp 145–198.
3. Gallistel CR (1990) *The Organization of Learning* (MIT Press, Cambridge, MA).
4. Pinker S (2002) *The Blank Slate* (Penguin, New York).
5. Wason P, Johnson-Laird P (1972) *Psychology of Reasoning* (Harvard Univ Press, Cambridge, MA).
6. Rips L (1994) *The Psychology of Proof* (MIT Press, Cambridge, MA).
7. Gigerenzer G, Murray D (1987) *Cognition as Intuitive Statistics* (Erlbaum, Hillsdale, NJ).
8. Rumelhart D, McClelland J, PDP Research Group (1986) *Parallel Distributed Processing* (MIT Press, Cambridge, MA).
9. Kahneman D (2003) A perspective on judgment and choice: Mapping bounded rationality. *Am Psychol* 58:697–720.
10. Gigerenzer G, Todd P, ABC Research Group (1999) *Simple Heuristics That Make Us Smart* (Oxford Univ Press, New York).
11. Gallistel CR, Gibbon J (2000) Time, rate, and conditioning. *Psychol Rev* 107:289–344.
12. Rode C, Cosmides L, Hell W, Tooby J (1999) When and why do people avoid unknown probabilities in decisions under uncertainty? *Cognition* 72:269–304.
13. Tooby J, Cosmides L, Barrett HC (2005) *The Innate Mind: Structure and Content*, eds Carruthers P, Laurence S, Stich S (Oxford Univ Press, New York), pp 305–337.
14. Cosmides L (1985) *Deduction or Darwinian Algorithms?* PhD dissertation (Harvard University).
15. Cosmides L (1989) The logic of social exchange. *Cognition* 31:187–276.
16. Cosmides L, Tooby J (2005) *The Handbook of Evolutionary Psychology*, ed Buss D (Wiley, Hoboken, NJ), pp 584–627.
17. Cosmides L, Tooby J (2008) *Moral Psychology*, ed Sinnott-Armstrong W (MIT Press, Cambridge, MA), pp 53–119.
18. Gigerenzer G, Hug K (1992) Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition* 43:127–171.
19. Stone V, Cosmides L, Tooby J, Kroll N, Knight R (2002) Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proc Natl Acad Sci USA* 99:11531–11536.
20. Ermer E, Guerin S, Cosmides L, Tooby J, Miller M (2006) Theory of mind broad and narrow: Reasoning about social exchange engages TOM areas, precautionary reasoning does not. *Soc Neurosci* 1:196–219.
21. Fiddick L, Spampinato M, Grafman J (2005) Social contracts and precautions activate different neurological systems: An fMRI investigation of deontic reasoning. *Neuroimage* 28:778–786.
22. Reis D, et al. (2007) Emotional intelligence predicts individual differences in social exchange reasoning. *Neuroimage* 35:1385–1391.
23. Sugiyama L, Tooby J, Cosmides L (2002) Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proc Natl Acad Sci USA* 99:11537–11542.
24. Fiddick L, Cosmides L, Tooby J (2000) No interpretation without representation. *Cognition* 77:1–79.
25. Cosmides L, Tooby J (1989) Evolutionary psychology and the generation of culture, part II. *Ethol Sociobiol* 10:51–97.
26. Gurven M (2004) To give and to give not. *Behav Brain Sci* 27:543–559.
27. Isaac G (1978) The food-sharing behavior of protohuman hominids. *Sci Am* 238: 90–108.
28. Tooby J, Cosmides L (1996) Friendship and the Banker's Paradox: Other pathways to the evolution of adaptations for altruism. *Evolution of Social Behaviour Patterns in Primates and Man. Proceedings of the British Academy*, eds Runciman WG, Maynard Smith J, Dunbar RIM, Vol 88, pp 119–143.
29. Trivers R (1971) The evolution of reciprocal altruism. *Q Rev Biol* 46:35–57.
30. Axelrod R (1984) *The Evolution of Cooperation* (Basic Books, New York).
31. Panchanathan K, Boyd R (2003) A tale of two defectors. *J Theor Biol* 224:115–126.
32. Wason P (1983) *Thinking and Reasoning*, ed Evans J (Routledge, London).
33. Harris P, Núñez M, Brett C (2001) Let's swap: Early understanding of social exchange by British and Nepali children. *Mem Cognit* 29:757–764.
34. Cheng P, Holyoak K (1985) Pragmatic reasoning schemas. *Cognit Psychol* 17:391–416.
35. Fodor J (2000) Why we are so good at catching cheaters. *Cognition* 75:29–32.
36. Manktelow K, Over D (1990) *Lines of Thinking*, eds Gilhooly K, Keane M, Logie R, Erdos G (Wiley, London), Vol 1, pp 153–164.
37. Manktelow K, Over D (1991) Social roles and utilities in reasoning with deontic conditionals. *Cognition* 39:85–105.
38. Sperber D, Cara F, Girotto V (1995) Relevance theory explains the selection task. *Cognition* 57:31–95.
39. Cheng P, Holyoak K (1989) On the natural selection of reasoning theories. *Cognition* 33:285–313.
40. Kirby K (1994) Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition* 51:1–28.
41. Oaksford M, Chater N (1994) A rational analysis of the selection task as optimal data selection. *Psychol Rev* 101:608–631.
42. Boyer P, Lienard P (2006) Why ritualized behavior? *Behav Brain Sci* 29:1–56.
43. Fiddick L (2004) Domains of deontic reasoning. *Q J Exp Psychol* 57:447–474.
44. Rumelhart D, Norman D (1981) *Cognitive Skills and Their Acquisition*, ed Anderson J (Erlbaum, Hillsdale, NJ).
45. Almor A, Sloman S (1996) Is deontic reasoning special? *Psychol Rev* 103:374–380.
46. Fiddick L, Rutherford M (2006) Looking for loss in all the wrong places: Loss-avoidance does not explain cheater detection. *Evol Hum Behav* 27:417–432.

Supporting Information
Adaptive specializations, social exchange, and the evolution of human intelligence

Leda Cosmides, H. Clark Barrett, & John Tooby

Subjects and Procedures. All subjects were students at universities in California. Experiments 1, 2, and 5 were conducted at Stanford University; Experiment 1-A was conducted at the University of California, Los Angeles (UCLA). Experiments 3 and 4 were conducted at the University of California, Santa Barbara (UCSB). Each subject was given a packet with an instruction sheet followed by one Wason selection task. They worked individually, seated at a quiet table with several other subjects, and were paid for their participation.

Experiment 1. Kalama benefit experiments.

Experiment 1 varied whether the action to be taken is a benefit to the potential violator (going out at night), neutral (staying home at night), or slightly unpleasant (taking out the garbage). The full text of the three conditions is presented here. Underlined portions indicate differences between conditions; no underlining appeared in the actual stimuli shown to subjects.

Condition 1: Benefit present (going out at night):

80% correct (20/25)

You are an anthropologist studying a tribe in the Kalama Islands. Members of this tribe always wear a small piece of volcanic rock tied around their ankle. These rocks come in an assortment of colors; people wear a different color of rock under different circumstances.

In this tribe, laws are made by a group of elders. One of the laws that the tribal elders made is: "If one is going out at night, then one must tie a small piece of red volcanic rock around one's ankle."

You are interested in whether the people of this tribe obey this law, so last night you watched what some of them did. The cards below represent four people who are member of the tribe. Each card represents one person. One side of a card tells whether the person went outside or stayed at home last night, and the other side of the card tells what color volcanic rock that person was wearing.

Indicate only those card(s) you definitely need to turn over **to see if any of these people have broken the law.**

A. went outside	B. red rock
C. stayed home	D. blue rock

Condition 2: No benefit version (staying home at night)

52% correct (13/25)

Everything identical except the following (underline indicates a change from *Condition 1*):

Rule: "If one is staying home at night, then one must tie a small piece of red volcanic rock around one's ankle."

To keep paragraph 3 parallel (with *P* mentioned before *not-P*), it said:
One side of a card tells whether the person stayed at home or went outside last night, and the other side of the card tells what color volcanic rock that person was wearing.

Condition 3: Garbage version

44% correct (11/25)

Everything identical except the following (underline indicates a change from *Condition 1*):

Paragraph 2: In this tribe, laws are made by a group of elders. Some of these laws involve chores. One of the laws that the tribal elders made is:
Rule: "If one is taking out the garbage, the one must tie a small piece of red volcanic rock around one's ankle."
Paragraph 3: You are interested in whether the people of this tribe obey this law, so one day you watched what some of them did. The cards below represent four people who are members of the tribe. Each card represents one person. Each of these people did one chore that day. One side of a card tells what chore the person did that day, and the other side of the card tells what color volcanic rock that person was wearing while doing the chore.
Cards: A. took out garbage B. red rock C: swept up embers D. blue rock

Note 1: Experiment 1, Benefits experiment. The same results obtain when it is made explicit that staying home is dreary compared to going out at night. In addition to the three conditions reported in the main text, we tested Condition 4 at Stanford at the same time. It was identical to the "staying home" problem of Condition 2 (rule 2), except for the insertion of the following paragraph after the first one, saying "The people of this tribe hate to stay home at night; they would much rather go outside and enjoy the nightly festival. Nonetheless, certain people are required to stay home at night to do dreary chores, which they find tedious and irritating." Performance on this problem was 48% correct (12/25), right between performance for rule 2 ("staying home" with no commentary: 52%) and rule 3 (the garbage chore problem: 44%).

Similarly, results from the Kalama requirement study at UCLA (Exp 1-A, below) show that performance on the going out at night problem is the same whether the story explicitly states that going out at night is a highly desirable activity (as in the UCLA study) or not (as in Condition 1 of the Stanford study). This is not surprising given the age and culture of college

students, and the use of the term "going out", when applied to nighttime activities, to indicate a date or other form of socializing/ entertainment.

For those interested in the role of negation in Wason tasks and in reasoning more generally, we note that performance on the going out at night problem at Stanford was the same, whether the problem used an "implicit not" (*not-Q* = blue rock, *Q* = red rock) as in Condition 1, or an "explicit not" (*not-Q* = did not wear a rock tied around ankle, *Q* = had rock tied around ankle) as in a fifth condition that was tested on the same Stanford population at the same time. This fifth condition was identical to Condition 1, except that it used explicit nots (and, therefore, omitted the two sentences "Members of this tribe always wear a small piece of volcanic rock tied around their ankle. These rocks come in an assortment of colors; people wear a different color of rock under different circumstances.") Performance on this "explicit not" version of the problem was 84% correct (21/25), virtually identical to the 80% correct found for the "implicit not" version (20/25). Many experiments have shown that use of an implicit not makes no difference when the rule clearly expresses a social contract, and detecting a violation corresponds to looking for cheaters.

Experiment 1-A. Kalama Requirement Study.

Experiment 1 varied whether the action regulated by a permission rule was a benefit or not; to count as a social contract, the rule must regulate access to a benefit. But what about the requirement term? Does it matter whether that is a cost or a benefit to the potential violator?

There have been some misunderstandings in the literature about this issue. Cheng and Holyoak (39), for example, thought that the hypothesis that there are social contract algorithms is refuted by evidence that violation detection remains high even when the requirement is not costly to the person who must satisfy it. But social contract theory makes no such prediction (see 14, 16, 25).

Social exchange is cooperation for mutual *benefit*; it is possible that one or both parties will suffer costs by providing benefits, but social exchange is not defined by the mutual imposition of costs. People enter into social exchange arrangements because they expect to *benefit* in some way. That providing a benefit to others can be costly was emphasized by Trivers (29) in his theory of the evolution of reciprocal altruism because his goal was to explain how adaptations for conferring benefits could be favored by selection *even* in cases where the conferrer suffers a cost by providing that benefit.

Of course, adaptations for reciprocation are easier to evolve when no costs are suffered, as in mutualism. Moreover, the ability to impose contingencies allows for increases in the welfare of both parties even when no cost is suffered by either. Suppose, e.g., that agent X is trying to choose between several personally beneficial actions—one of which is also beneficial to agent Y. Y can influence X's choice through social exchange: If Y controls access to something further that X wants, Y can make it available to X contingent on X choosing the option that benefits Y (25).

In social exchange, an agent imposes a requirement because doing so creates a situation that agent wants. Requiring that people be over 21 before they drink alcohol creates safer streets, which is why lawmakers made access to alcohol contingent on this requirement; yet the

state of being over or under 21 is not, by itself, a cost or a benefit to the lawmakers. A mother may tell her child that she can have a playdate if she gets her homework done—even though doing the homework benefits the distractible child while stressing the mother, and even though the mother, who needs some quiet time to get work done, will be worse off if the playdate does *not* happen. She makes the offer because, if the child follows through, it creates a situation that the mother prefers: Playdate along with completed homework.

In Experiment 1-A we tested the prediction that violation detection will remain high whether the requirement is a cost or benefit to the person who must satisfy it. We also explored whether it matters whether the requirement itself is beneficial to the agent(s) who imposed it. This study was conducted at UCLA, in a large introductory anthropology class. Each subject filled out a packet consisting of an instruction sheet followed by one Wason task.

The problems were versions of Condition 1 of Experiment 1, the Kalama "going out at night" problem. In all of them, the story specified that Kalama teenagers (the potential rule violators) consider going out at night to be a benefit. It explained that the elders have made a law that applies to teenagers: "If one is going out at night, then one must tie a diamond around one's ankle" (the rock was a diamond in some versions, a pebble in others, just to see if the real-world difference in their value affected performance (it did not)).

What varied across problems was whether a teen wearing a rock is considered a cost or benefit by (i) the teens who are wearing them, and/or (ii) the elders who imposed this requirement. Here is the Elders love/ Teens hate Diamond version, with the key paragraphs underlined:

You are an anthropologist studying a tribe in the Kalama Islands. In this tribe, laws are made by a group of elders.

Kalama teen-agers are very fun-loving. They like to go out at night, to party and visit their friends.

<u>Kalama teen-agers hate to wear a diamond tied around their ankle; among Kalama teen-agers, it is considered ridiculous and embarrassing</u>.

<u>But the tribal elders love it when teen-agers wear a diamond tied around their ankle. They think it looks beautiful, and like it when people from other tribes admire their children.</u>

The tribal elders have made a law that applies to teen-agers:
"If one is going out at night, then one must tie a diamond around one's ankle."

You are interested in whether the teen-agers obey this law, so last night you watched what some of them did. The cards below represent four teen-agers who are members of this tribe. Each card represents one teen-ager. One side of a card tells whether or not the teen-ager went out last night, and the other side of the card tells whether or not that teen-ager had a diamond tied around his or her ankle.

Indicate only those card(s) you definitely need to turn over **to see if any of these teen-agers have broken the law.**

Cards: A. went out B. had diamond tied around ankle C. did not go out D. did not have diamond tied around ankle

For conditions in which the teens like the requirement, the first underlined paragraph said this instead: Kalama teen-agers love to wear a diamond tied around their ankle; among Kalama teen-agers, it is considered very hip.

For conditions in which the elders do not like the requirement, but have imposed it anyway (presumably to create a situation they desire), the second underlined paragraph said this instead: But the tribal elders hate it when teen-agers wear a diamond tied around their ankle. They think it looks cheap and degrading, and that the teen-agers might lose them.

In conditions where the tribal elders and teens are in agreement, this second paragraph started "The tribal elders also [hate / love] it when…"

There were conditions where both of the underlined paragraphs were missing, providing no information about whether teens or elders like it when teens wear a diamond (or pebble); there was also a non-social contract permission rule as a point of comparison (the alphanumeric rule, see Exp 3, condition 5, below).

Exp 1-A Kalama requirement study	% correct		Attitude toward teens wearing ankle rock	
Condition		# correct /N	Elders	Teens
1. Pebble (no info)	74.3%	26/35	?	?
2. Diamond (no info)	75.7%	28/37	?	?
3. Diamond EL-TH	70.6%	24/34	love it	hate it
4. Diamond EL-TL	71.4%	26/36	love it	love it
5. Diamond EH-TH	69.7%	23/33	hate it	hate it
6. Pebble EH-TH	73.5%	25/34	hate it	hate it
7. Diamond EH-TL	**50.0%**	**17/34**	**hate it**	**love it**
8. Abstract deontic	9.7%	3/31	--	--

The results were quite clear. Performance was equally high for conditions 1-6. When there was no information, it did not matter if the rock was a pebble or a diamond (conditions 1 and 2). When it was a diamond, performance was just as high when the elders like teens wearing the rock and teens hate it (condition 3) as when elders and teens both like it (condition 4) or both hate it (condition 5; also replicated with the pebble in condition 6).

The only drop in performance was for condition 7: the condition in which elders hate the requirement and teens like it ($0.012 \leq P\text{s} \leq 0.05$, $0.20 \leq phi \leq .27$). Condition 7 presents the most puzzling set of motivations from a social exchange viewpoint: one wonders why the elders are requiring something that they hate and that the teens are motivated to do anyway. It took the conjunction of these two conditions to make a dent in performance. Nevertheless, this rule is a still a social contract, regulating access to a benefit, and it elicited higher performance than the abstract deontic rule tested ($P = 0.0002$, phi = 0.44).

Experiment 2. The Sears problem.

The Sears problem was developed by Roy D'Andrade, a cultural anthropologist at the University of California, San Diego, during the 1970s. The fact that about 70% of subjects answered correctly (*P & not-Q*) was reported in (44). The authors thought performance was so much higher than on other (indicative) Wason tasks because Sears was a culturally familiar context. Since then, it has become clear that familiarity alone is not sufficient to elicit high levels of violation detection (14, 15, 32); familiarity is not necessary either—highly unfamiliar rules can elicit high performance if they are social contracts or precaution rules (14, 15, 39, 43; see also Experiment 1 above, and the unfamiliar precautionary garbage rule, reported below).

Condition 1. Benefit present (high versus low value good).

72% correct (18/25)

The wording of this problem is identical to that of the original Sears problem created by D'Andrade (reported in 44). For this reason, the final question refers to seeing whether the sales clerks "followed" the regulation, rather than seeing whether anyone had violated the rule.

You work as an assistant at Sears. You have the job of checking sales receipts to make sure that any sale over $30 has been approved by the section manager (This is a rule of the store)

The cards below have information about four sales receipts. Each card represents one sales receipt. One side of a card tells how much the sale was for, and the other side of the card tells whether or not that sales receipt had been signed by the section manager.

Which of the cards would you have to turn over to make sure the sales clerks had followed the regulation?

Cards: A. $70 B. signed C. $15 D. not signed

Condition 2. Inventory forms, signature required.

48% correct (12/25)

You work as an assistant at Sears. Each department at Sears (Menswear, Sportswear, Ladies Shoes, etc.) has a different color inventory form to fill out. You have the job of checking inventory forms that the department clerks have filled out to make sure that any blue inventory form has been signed by the section manager. (This is a rule of the store.)

The cards below have information about four inventory forms that have been filled out by department clerks. Each card represents one inventory form. One side of a card tells what color the inventory form is, and the other side of the card tells whether or not that inventory form has been signed by the section manager.

Which of the cards would you have to turn over to make sure the department clerks had followed the regulation?

Cards: A. blue B. signed C. white D. not signed

Condition 3. Inventory forms, filing requirements.

32% correct (8/25)

You work as an assistant at Sears. Each department at Sears (Menswear, Sportswear, Ladies Shoes, etc.) has a different color inventory form to fill out. Filled out inventory forms are to be filed in various bins. You have the job of checking inventory forms that the department clerks have filled out to make sure that any blue inventory form has been filed in the metal bin. (This is a rule of the store.)

The cards below have information about four inventory forms that have been filled out by department clerks. Each card represents one inventory form. One side of a card tells what color the inventory form is, and the other side of the card tells what kind of bin it was filed in.

Which of the cards would you have to turn over to make sure the department clerks had followed the regulation?

Cards: A. blue B. metal bin C. white D. wood bin

Experiment 3. School problems, Innocent mistake Series 1.

Condition 1. Social contract rule, Cheater detection version.

68% correct (23/34)

You supervise four <u>women</u> who volunteered to help out at the local Board of Education. When you came into work today, you found the place a-buzz with rumor and innuendo. Your volunteers were supposed to follow certain rules for assigning students from various towns to the appropriate school district. Each volunteer is the mother of a teenager who is about to enter high school, and each processed her own child's documents. So now rumors are flying that your volunteers cheated on the rules when it came to assigning their own children to a school.

Here is the situation:

Students are to be assigned either to Dover High School, which is located in Dover City, or to Hanover High School, which is located in the town of Hanover. Dover High is a great school with an excellent record for getting students placed in good colleges. In contrast, Hanover High is a mediocre school with poor teachers and decrepit facilities.

The reason the schools are so different is how willing the parents of each community are to financially support their schools through taxes. Although both communities are equally prosperous, the parents in Dover City have always cared about the quality of their schools, including Dover High, and have been willing to pay for it. In contrast, the parents in the neighboring towns of Hanover and Belmont have never wanted to spend the money, and have opposed any taxes to improve Hanover High.

The Board of Education took these factors into account when it created rules to determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Dover High School, then that student must live in Dover City."

Your volunteers were supposed to follow this rule when processing *all* student documents—including the documents of their own children! You must find out if the rumors are true: did any of your volunteers cheat on this rule when it came to processing their own children's documents?

The cards below have information about the documents of the four volunteers' children. Each card represents the child of one volunteer. One side of a card tells what school the volunteer assigned her son or daughter to, and the other side of the card tells what town that student lives in.

Most parents want their children to get the best education possible, however, some are not willing to pay for it. It is easy to imagine that your volunteers, being ambitious mothers, might have been tempted to cheat on the rule. Indicate only those card(s) you definitely need to turn over **to see if the documents of any of these students violate the rule.**

A. Dover High School	B. town of Hanover
C. Hanover High School	D. Dover City

Condition 2. Social contract rule, Sabotage version.

45% correct (15/33)

You supervise four women who volunteered to help out at the local Board of Education. When you came into work today, you found the place a-buzz with rumor and innuendo. Your volunteers were supposed to follow certain rules for assigning students from various towns to the appropriate school district. They are mad at you, however, because you fired their best friend last week. So how rumors are flying that your volunteers have purposely violated the rules when it came to assigning students to a school, in an attempt to create chaos and to make you look like an incompetent supervisor in the eyes of your boss.

Here is the situation:

Students are to be assigned either to Dover High School, which is located in Dover City, or to Hanover High School, which is located in the town of Hanover. Dover High is a great school with an excellent record for getting students placed in good colleges. In contrast, Hanover High is a mediocre school with poor teachers and decrepit facilities.

The reason the schools are so different is how willing the parents of each community are to financially support their schools through taxes. Although both communities are equally prosperous, the parents in Dover City have always cared about the quality of their schools, including Dover High, and have been willing to pay for it. In contrast, the parents in the

neighboring towns of Hanover and Belmont have never wanted to spend the money, and have opposed any taxes to improve Hanover High.

Most parents want their children to get the best education possible, however, some are not willing to pay for it.

The Board of Education took these factors into account when it created rules to determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Dover High School, then that student must live in Dover City."

Your volunteers were supposed to follow this rule when processing student documents. You must find out if they the rumors are true: did any of your volunteers violate this rule when it came to processing student documents?

The cards below have information about the documents of four students. Each card represents one student. One side of a card tells what school the volunteer assigned the student to, and the other side of the card tells what town the student lives in.

Your volunteers are very angry. It is easy to imagine that they might want to violate the rule; the chaos this would create would make you look incompetent and get you in trouble with your boss. Indicate only those card(s) you definitely need to turn over **to see if the documents of any of these students violate the rule.**

Condition 3. Social contract rule, Innocent mistake version.

27% correct (9/33)

You supervise four elderly ladies who volunteered to help out at the local Board of Education. Your volunteers were supposed to follow certain rules for assigning students from various towns to the appropriate school district. They are all sweet little old ladies, but they have become rather absent-minded lately. So you are worried that they may have inadvertently made some mistakes when it came to assigning students to a school.

Here is the situation:

Students are to be assigned either to Dover High School, which is located in Dover City, or to Hanover High School, which is located in the town of Hanover. Dover High is a great school with an excellent record for getting students placed in good colleges. In contrast, Hanover High is a mediocre school with poor teachers and decrepit facilities.

The reason the schools are so different is how willing the parents of each community are to financially support their schools through taxes. Although both communities are equally prosperous, the parents in Dover City have always cared about the quality of their schools, including Dover High, and have been willing to pay for it. In contrast, the parents in the neighboring towns of Hanover and Belmont have never wanted to spend the money, and have opposed any taxes to improve Hanover High.

Most parents want their children to get the best education possible, however, some are not willing to pay for it.

The Board of Education took these factors into account when it created rules to determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Dover High School, then that student must live in Dover City."

Your volunteers were supposed to follow this rule when processing student documents. You must find out if they made any mistakes: did any of your volunteers inadvertently violate this rule when they were processing student documents?

The cards below have information about the documents of four students. Each card represents one student. One side of a card tells what school the volunteer assigned the student to, and the other side of the card tells what town the student lives in.

Your volunteers are very nice and helpful ladies who try to follow the rules, but they have been getting so absent-minded lately that it is easy to imagine that they might have made some mistakes. Indicate only those card(s) you definitely need to turn over **to see if the documents of any of these students violate the rule.**

Condition 4. Permission rule that is NOT a social contract rule: School rule.

10% correct (3/29) Although the rule is identical, the story has been changed in condition 4, removing the elements that made it a social contract in conditions 1-3. Although performance in condition 3—a social contract with mistakes—it was even lower for this rule, which is not a social contract ($P = 0.046$)

The secretary you replaced at the local Board of Education may have made some mistakes when she processed student documents. It is important that certain rules for assigning students from various towns to the appropriate school district be followed, because the population statistics they provide allow the Board of Education to decide how many teachers need to be assigned to each school. If these rules are not followed, some schools would end up with too many teachers, and other schools with too few.

Students are to be assigned either to Dover High School or to Hanover High School.

Some students live in the town of Dover City, some live in Hanover, and some live in Belmont. There are rules that determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Dover High School, then that student must live in Dover City."

Shortly before she retired, the secretary you replaced was supposed to sort through the documents that specify what town the students live in, and make school assignments according to this rule. She was a sweet little old lady who had become rather absent-minded, and who often made mistakes when categorizing student documents.

The cards below have information about the documents of four students. Each card represents one student. One side of a card tells what school the retired secretary assigned the student to, and the other side of the card tells what town the student lives in.

You suspect the retired secretary may have inadvertently categorized some of the students' documents incorrectly, so you decide to see for yourself whether she ever violated the rule. Indicate only those card(s) you definitely need to turn over **to see if the documents of any of these students violate the rule.**

Condition 5. Permission rule that is NOT a social contract: Alphanumeric rule.

6% correct (2/31)

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

"If a student has a 'D' rating, then that student's documents must be marked code '3'".

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

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

Cards read: A. **D** B. **3** C. **F** D. **7**

Experiment 4. Parametric study of benefits, intentions, and ability.

In this study, we systematically varied whether the potential for violators to gain the *benefit* regulated by the rule was present or absent (B+, B-), whether the *intention* to violate was present or absent (I+, I-), and whether the situation provided them with the *ability* to violate the rule (A+, A-). In ability absent conditions (A-), the scenario stated that students were identified by code numbers to prevent volunteers from knowing whose documents they were processing.

<u>Underlining</u> denotes portions of stimuli that varied across conditions. Underlining did not appear in actual stimuli.

Benefit Present conditions:

Intention present: B+I+A+; B+I+A-

BIA (Benefit +, Intent +, Ability +)

64% correct (49/77)

You supervise four women who volunteered to help out at the local Board of Education. Your volunteers were supposed to follow certain rules for assigning students from various towns

to the appropriate school district. Each volunteer is the mother of a teenager who is about to enter high school, and each processed her own child's documents. You overheard that some of your volunteers intended to break the rules when it came to assigning their own children to a school. Here is the situation:

Students are to be assigned either to Grover High School, which is located in Grover City, or to Hanover High School, which is located in the town of Hanover. Grover High is a great school with an excellent record for getting students placed in good colleges. In contrast, Hanover High is a mediocre school with poor teachers and decrepit facilities.

The reason the schools are so different is how willing the parents of each community are to financially support their schools through taxes. Although both communities are equally prosperous, the parents in Grover City have always cared about the quality of their schools, including Grover High, and have been willing to pay for it. In contrast, the parents in the neighboring towns of Hanover and Belmont have never wanted to spend the money, and have opposed any taxes to improve Hanover High.

The Board of Education took these factors into account when it created rules to determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Grover High School, then that student must live in Grover City."

Your volunteers were supposed to follow this rule when processing all student documents, including the documents of their own children. However, you overheard that some of your volunteers intended to break this rule. You must find out if any of your volunteers broke the rule.

The cards below have information about the documents of children that the volunteers processed. Each card represents the child of one volunteer. One side of a card tells what school the volunteer assigned her son or daughter to, and the other side of the card tells what town the student lives in.

It could benefit a volunteer if her child were assigned to a good school that she did not pay for. You overheard that some of your volunteers intended to break the rule. Please indicate only those card(s) you need to turn over **to see if the documents of any of these students violate the rule.**

A. Grover High School B. Grover City C. Hanover High School D. town of Hanover

BI (Benefit +, Intent +, Ability -)

46% correct (18/39)

You supervise four women who volunteered to help out at the local Board of Education. Your volunteers were supposed to follow certain rules for assigning students from various towns to the appropriate school district. Each volunteer is the mother of a teenager who is about to enter high school. The students were identified by code numbers so that parents could not know which documents were those of her own child. Nevertheless, you overheard that some of your

volunteers intended to try to break the rules when it came to assigning children to schools. Here is the situation:

Students are to be assigned either to Grover High School, which is located in Grover City, or to Hanover High School, which is located in the town of Hanover. Grover High is a great school with an excellent record for getting students placed in good colleges. In contrast, Hanover High is a mediocre school with poor teachers and decrepit facilities.

The reason the schools are so different is how willing the parents of each community are to financially support their schools through taxes. Although both communities are equally prosperous, the parents in Grover City have always cared about the quality of their schools, including Grover High, and have been willing to pay for it. In contrast, the parents in the neighboring towns of Hanover and Belmont have never wanted to spend the money, and have opposed any taxes to improve Hanover High.

The Board of Education took these factors into account when it created rules to determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Grover High School, then that student must live in Grover City."

Your volunteers were supposed to follow this rule when processing all student documents, including the documents of their own children. Student documents were identified by code numbers only, not names, to ensure this. Nevertheless, you overheard that some of your volunteers intended to break this rule. You must find out if any of your volunteers broke the rule.

The cards below have information about the documents of children that the volunteers processed. Each card represents the child of one volunteer. One side of a card tells what school the volunteer assigned the student to, and the other side of the card tells what town the student lives in.

It could benefit a volunteer if her child were assigned to a good school that she did not pay for. Even though a volunteer could not know which documents were those of her child, you overheard that some of your volunteers intended to break the rule. Please indicate only those card(s) you need to turn over **to see if the documents of any of these students violate the rule.**

Benefit Present conditions:

Intention absent: B+I-A+; B+I-A-

BA (Benefit +, Intent -, Ability +)

41% correct (16/39)

You supervise four women who volunteered to help out at the local Board of Education. Your volunteers were supposed to follow certain rules for assigning students from various towns to the appropriate school district. Each volunteer is the mother of a teenager who is about to enter high school, and each processed her own child's documents. You believe your volunteers are honest, but you suspect that they may have made some innocent mistakes: they may have broken the rules for assigning each child to a particular school. Here is the situation:

Students are to be assigned either to Grover High School, which is located in Grover City, or to Hanover High School, which is located in the town of Hanover. Grover High is a great school with an excellent record for getting students placed in good colleges. In contrast, Hanover High is a mediocre school with poor teachers and decrepit facilities.

The reason the schools are so different is how willing the parents of each community are to financially support their schools through taxes. Although both communities are equally prosperous, the parents in Grover City have always cared about the quality of their schools, including Grover High, and have been willing to pay for it. In contrast, the parents in the neighboring towns of Hanover and Belmont have never wanted to spend the money, and have opposed any taxes to improve Hanover High.

The Board of Education took these factors into account when it created rules to determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Grover High School, then that student must live in Grover City."

Your volunteers were supposed to follow this rule when processing all student documents, <u>including the documents of their own children. You believe your volunteers are honest, but you suspect that they may have made some innocent mistakes; they may have broken this rule</u>. You must find out if any of your volunteers broke the rule.

The cards below have information about the documents of children that the volunteers processed. <u>Each card represents one child.</u> One side of a card tells what school the volunteer assigned <u>the child</u> to, and the other side of the card tells what town the student lives in.

<u>Although it could benefit a volunteer if her child were assigned to a good school that she did not pay for, you suspect your volunteers may have broken the rule by mistake.</u> Please indicate only those card(s) you need to turn over **to see if the documents of any of these students violate the rule.**

B (Benefit +, Intent -, Ability -)

26% correct (10/39)

You supervise four women who volunteered to help out at the local Board of Education. Your volunteers were supposed to follow certain rules for assigning students from various towns to the appropriate school district. <u>Each volunteer is the mother of a teenager who is about to enter high school. The students were identified by code numbers so that parents could not know which documents were those of her own child. You believe your volunteers are honest, you suspect that they may have made some innocent mistakes: they may have broken the rules for assigning each child to a particular school.</u> Here is the situation:

Students are to be assigned either to Grover High School, which is located in Grover City, or to Hanover High School, which is located in the town of Hanover. Grover High is a great school with an excellent record for getting students placed in good colleges. In contrast, Hanover High is a mediocre school with poor teachers and decrepit facilities.

The reason the schools are so different is how willing the parents of each community are to financially support their schools through taxes. Although both communities are equally prosperous, the parents in Grover City have always cared about the quality of their schools, including Grover High, and have been willing to pay for it. In contrast, the parents in the neighboring towns of Hanover and Belmont have never wanted to spend the money, and have opposed any taxes to improve Hanover High.

The Board of Education took these factors into account when it created rules to determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Grover High School, then that student must live in Grover City."

Your volunteers were supposed to follow this rule when processing all student documents, including the documents of their own children. Student documents were identified by code numbers only, not names, to ensure this. You believe your volunteers are honest, but you suspect that they may have made some innocent mistakes; they may have broken this rule. You must find out if any of your volunteers broke the rule.

The cards below have information about the documents of children that the volunteers processed. Each card represents the child of one volunteer. One side of a card tells what school the volunteer assigned the student to, and the other side of the card tells what town the student lives in.

It could benefit a volunteer if her child were assigned to a good school that she did not pay for. Even though a volunteer could not know which documents were those of her child, you suspect your volunteers may have broken the rule by mistake Please indicate only those card(s) you need to turn over **to see if the documents of any of these students violate the rule.**

Benefit absent conditions:

Intention present: B-I+A+, B-I+A-

IA (Benefit -, Intent +, Ability +)

44% correct (17/39)

You supervise four women who volunteered to help out at the local Board of Education. Your volunteers were supposed to follow certain rules for assigning students from various towns to the appropriate school district None of your volunteers have children in school and so could not benefit from their positions, but you overheard that some of them may have mischievously intended to break the rules when assigning children to schools. Here is the situation:

Students are to be assigned either to Grover High School, which is located in Grover City, or to Hanover High School, which is located in the town of Hanover. Grover High is a great school with an excellent record for getting students placed in good colleges. In contrast, Hanover High is a mediocre school with poor teachers and decrepit facilities.

The reason the schools are so different is how willing the parents of each community are to financially support their schools through taxes. Although both communities are equally

prosperous, the parents in Grover City have always cared about the quality of their schools, including Grover High, and have been willing to pay for it. In contrast, the parents in the neighboring towns of Hanover and Belmont have never wanted to spend the money, and have opposed any taxes to improve Hanover High.

The Board of Education took these factors into account when it created rules to determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Grover High School, then that student must live in Grover City."

Your volunteers were supposed to follow this rule when processing all students. However, you overheard that some of your volunteers mischievously intended to break this rule. You must find out if any of your volunteers broke the rule.

The cards below have information about the documents of children that the volunteers processed. Each card represents one student. One side of a card tells what school the volunteer assigned the student to, and the other side of the card tells what town the student lives in.

Although your volunteers do not have children in school and so could not benefit by mis-assigning students, you overheard that some of your volunteers intended to break the rule. Please indicate only those card(s) you need to turn over **to see if the documents of any of these students violate the rule.**

I (Benefit -, Intent +, Ability -)

28% correct (11/39)

You supervise four women who volunteered to help out at the local Board of Education. Your volunteers were supposed to follow certain rules for assigning students from various towns to the appropriate school district. None of your volunteers have children in school and so could not benefit from their positions. Furthermore, the students were identified by code numbers so that volunteers could not know which documents were those of which child. Nevertheless, you overheard that some of your volunteers intended to break the rules when assigning children to schools. Here is the situation:

Students are to be assigned either to Grover High School, which is located in Grover City, or to Hanover High School, which is located in the town of Hanover. Grover High is a great school with an excellent record for getting students placed in good colleges. In contrast, Hanover High is a mediocre school with poor teachers and decrepit facilities.

The reason the schools are so different is how willing the parents of each community are to financially support their schools through taxes. Although both communities are equally prosperous, the parents in Grover City have always cared about the quality of their schools, including Grover High, and have been willing to pay for it. In contrast, the parents in the neighboring towns of Hanover and Belmont have never wanted to spend the money, and have opposed any taxes to improve Hanover High.

The Board of Education took these factors into account when it created rules to determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Grover High School, then that student must live in Grover City."

Your volunteers were supposed to follow this rule when processing all student documents. Student documents were identified by code numbers only, not names, to ensure that volunteers would not know whose documents they were processing. However, you overheard that some of your volunteers intended to break this rule. You must find out if any of your volunteers broke the rule.

The cards below have information about the documents of children that the volunteers processed. Each card represents one student. One side of a card tells what school the volunteer assigned the student to, and the other side of the card tells what town the student lives in.

Although your volunteers do not have children in school and so could not benefit by mis-assigning students, and although student documents were identified by code numbers, you overheard that some of your volunteers intended to break the rule. Please indicate only those card(s) you need to turn over **to see if the documents of any of these students violate the rule.**

Benefit absent conditions:

Intention absent: B-I-A+, B-I-A-

A (Benefit -,Intent -, Ability +)

23% correct (9/39)

You supervise four women who volunteered to help out at the local Board of Education. Your volunteers were supposed to follow certain rules for assigning students from various towns to the appropriate school district None of your volunteers have children in school and so could not benefit from their positions. You believe your volunteers are honest, but you suspect that they may have made some innocent mistakes: they may have broken the rules for assigning each child to a particular school. Here is the situation:

Students are to be assigned either to Grover High School, which is located in Grover City, or to Hanover High School, which is located in the town of Hanover. Grover High is a great school with an excellent record for getting students placed in good colleges. In contrast, Hanover High is a mediocre school with poor teachers and decrepit facilities.

The reason the schools are so different is how willing the parents of each community are to financially support their schools through taxes. Although both communities are equally prosperous, the parents in Grover City have always cared about the quality of their schools, including Grover High, and have been willing to pay for it. In contrast, the parents in the neighboring towns of Hanover and Belmont have never wanted to spend the money, and have opposed any taxes to improve Hanover High.

The Board of Education took these factors into account when it created rules to determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Grover High School, then that student must live in Grover City."

Your volunteers were supposed to follow this rule when processing all students. You believe your volunteers are honest, but you suspect that they may have made some innocent mistakes; they may have broken this rule. You must find out if any of your volunteers broke the rule.

The cards below have information about the documents of children that the volunteers processed. Each card represents one student. One side of a card tells what school the volunteer assigned the student to, and the other side of the card tells what town the student lives in.

Although your volunteers do not have children in school and so could not benefit by mis-assigning students, you suspect your volunteers may have broken the rule by mistake. Please indicate only those card(s) you need to turn over **to see if the documents of any of these students violate the rule.**

None (Benefit -, Intent -, Ability –)

26% correct (10/39) This condition was added to the study later, for completeness. The subjects were drawn from the same UCSB student population; we confirmed that the population sampled at this later time was the same as that sampled earlier by giving the BIA condition to 39 more subjects at the same time we collected the data for the "none" condition. The BIA percent correct was virtually identical to that found earlier (24/38 and 25/39; resulting in an overall percent correct of 49/77 for the BIA condition).

You supervise four women who volunteered to help out at the local Board of Education. Your volunteers were supposed to follow certain rules for assigning students from various towns to the appropriate school district. None of your volunteers have children in school and so could not benefit from their positions. Furthermore, the students were identified by code numbers so that volunteers could not know which documents were those of which child. You believe your volunteers are honest, but you suspect that they may have made some innocent mistakes: they may have broken the rules for assigning each child to a particular school. Here is the situation:

Students are to be assigned either to Grover High School, which is located in Grover City, or to Hanover High School, which is located in the town of Hanover. Grover High is a great school with an excellent record for getting students placed in good colleges. In contrast, Hanover High is a mediocre school with poor teachers and decrepit facilities.

The reason the schools are so different is how willing the parents of each community are to financially support their schools through taxes. Although both communities are equally prosperous, the parents in Grover City have always cared about the quality of their schools, including Grover High, and have been willing to pay for it. In contrast, the parents in the neighboring towns of Hanover and Belmont have never wanted to spend the money, and have opposed any taxes to improve Hanover High.

The Board of Education took these factors into account when it created rules to determine which school a student is to be assigned to; the most important of these rules is:

"If a student is to be assigned to Grover High School, then that student must live in Grover City."

Your volunteers were supposed to follow this rule when processing all student documents. Student documents were identified by code numbers only, not names, to ensure that volunteers would not know whose documents they were processing. You believe your volunteers are honest, but you suspect that they may have made some innocent mistakes; they may have broken this rule. You must find out if any of your volunteers broke the rule.

The cards below have information about the documents of children that the volunteers processed. Each card represents one student. One side of a card tells what school the volunteer assigned the student to, and the other side of the card tells what town the student lives in.

Although your volunteers do not have children in school and so could not benefit by mis-assigning students, and although student documents were identified by code numbers, you suspect your volunteers may have broken the rule by mistake. Please indicate only those card(s) you need to turn over **to see if the documents of any of these students violate the rule.**

Note 2. The effects of removing the intention to violate the rule (*intention* absent) and introducing measures to prevent cheating (*ability* absent) are all the more remarkable when one considers the *benefit present* problems alone. Although none of them said that the rule has been violated when a student who lives in Hanover is assigned to Grover High, the four problems in this series each said "It could benefit a volunteer if her child were assigned to a good school that she did not pay for." This is as close as one can come to explaining what counts as violating the rule without saying it outright. Nevertheless, there was a significant drop in violation detection when either the intention or the ability to violate (but not both) were removed—from 64% (BIA) to 41% (BA, $P = 0.01$) and 46% (BI, $P = 0.036$)—and removing both factors dropped performance to 26% (B).

Note 3. In Experiment 4, the rule is created by members of the Board of Education. To elicit good performance on social contract rules, the story does not need to specify what specific goal the agent who is restricting access to the benefit is trying to achieve by doing so (e.g., Exp 1, the going out problem). The Board's intent may have been to offer better service to those who pay more, or to incentivize tax support for local schools from the prosperous citizens of the other towns. But whatever their goal was, it is thwarted when the rule is violated. Failure to comply results in them being cheated, even if it was not caused by a cheater.

Experiment 5

Precautionary rule, Garbage version.

72% correct, 18/25 Stanford students. Performance on this problem was significantly higher than on the very similar garbage problem tested in Experiment 1.

The high level of performance found on this problem shows two things. First, the garbage rule (rule 3) is capable of eliciting strong violation detection—when it is embedded in a

precautionary context. Second, it shows that, when a rule is precautionary, it does not need to regulate access to a *benefit* to elicit good violation detection. This is another way in which reasoning about precautionary rules differs from reasoning about social contracts.

A tribe in the Kalama Islands believes that vicious spirits attack people when they are taking out the garbage, but that these spirits do not attack people at any other time. These people also believe that fastening a small piece of volcanic rock (which is abundant and free on the islands) around one's ankle will make one invisible to the spirits and thus safe from them. Tribespeople therefore have the following rule:
"If one is taking out the garbage, then one must tie a small piece of volcanic rock around one's ankle."

You are an anthropologist studying this tribe. You are interested in whether the people of this tribe follow this rule, so one day you watched what some of them did. The cards below represent four tribespeople who you observed that day. Each card represents one person. One side of a card tells whether or not the person took out the garbage, and the other side of the card tells whether or not that person had a piece of volcanic rock tied around his ankle.

Indicate only those card(s) you definitely need to turn over **to see if any of these people have broken the rule.**

A. took out garbage B. had rock tied around ankle C. did not take out garbage D. did not have rock tied around ankle

In this problem, not-Q is an "explicit not" (*not-Q* = not wearing a rock, *Q* = wearing a rock), whereas it is an "implicit not" in Exp 1 (*not-Q* = wearing a blue rock, *Q* = wearing a red rock). Use of an implicit not cannot explain the poorer performance of this rule in Exp 1, for two reasons. First, the matching problem about going out at night used the exact same implicit not as the garbage rule, yet elicited the correct response from 80% of subjects. Second, that 80% correct was not different from the 84% correct (21/25) obtained when we ran a version of the going out at night problem that used an explicit not. Many experiments have shown that use of an implicit not makes no difference when the rule clearly expresses a social contract.

Exp 1-A Kalama requirement study	% correct		Attitude toward teens wearing ankle rock	
Condition		# correct /N	Elders	Teens
1. Pebble (no info)	74.3%	26/35	?	?
2. Diamond (no info)	75.7%	28/37	?	?
3. Diamond EL-TH	70.6%	24/34	love it	hate it
4. Diamond EL-TL	71.4%	26/36	love it	love it
5. Diamond EH-TH	69.7%	23/33	hate it	hate it
6. Pebble EH-TH	73.5%	25/34	hate it	hate it
7. Diamond EH-TL	**50.0%**	**17/34**	**hate it**	**love it**
8. Abstract deontic	9.7%	3/31	--	--