Title

Theory of mind broad and narrow: Reasoning about social exchange engages ToM areas, precautionary reasoning does not

Journal

SOCIAL NEUROSCIENCE, 1

ISSN

1747-0919

Authors

Ermer, Elsa
Guerin, Scoft A.
Cosmides, Leda
et al.

Publication Date

2006

Peer reviewed

Theory of mind broad and narrow:

Reasoning about social exchange engages TOM areas, precautionary reasoning does not

Elsa Ermer[1,2], Scott A. Guerin[1], Leda Cosmides[1,2], John Tooby[2], Michael B. Miller[1]

1. Department of Psychology, University of California, Santa Barbara
2. Center for Evolutionary Psychology, University of California, Santa Barbara

Short Title: Social exchange reasoning engages TOM

Social Neuroscience, 1, 196-219.
Special Issue on Neural Correlates of Theory of Mind
Rebecca Saxe and Simon Baron-Cohen, Issue editors

Corresponding author:
Elsa Ermer
Department of Psychology
University of California
Santa Barbara, CA 93106-9660
ermer@psych.ucsb.edu,
Phone: 805-452-4988
Fax: 805-965-1163

Keywords: Theory of Mind, Social exchange, Deontic reasoning, Cheater detection, Neuroimaging

Abstract

Baron-Cohen (1995) proposed that the theory of mind (TOM) inference system evolved to promote strategic social interaction. Social exchange—a form of cooperation for mutual benefit—involves strategic social interaction and requires TOM inferences about the contents of other individual's mental states, especially their desires, goals, and intentions. There are behavioral and neuropsychological dissociations between reasoning about social exchange and reasoning about equivalent problems tapping other, more general, content domains. It has therefore been proposed that social exchange behavior is regulated by *social contract algorithms*: a domain-specific inference system that is functionally specialized for reasoning about social exchange. We report an fMRI study using the Wason selection task that provides further support for this hypothesis. Precautionary rules share so many properties with social exchange rules—they are conditional, deontic, and involve subjective utilities—that most reasoning theories claim they are processed by the same neurocomputational machinery. Nevertheless, neuroimaging shows that reasoning about social exchange activates brain areas not activated by reasoning about precautionary rules, and vice versa. As predicted, neural correlates of theory of mind (anterior and posterior temporal cortex) were activated when subjects interpreted social exchange rules, but not precautionary rules (where TOM inferences are unnecessary). We argue that the interaction between TOM and social contract algorithms can be reciprocal: social contract algorithms requires TOM inferences, but their functional logic also allows TOM inferences to be made. By considering interactions between TOM in the narrower sense (belief-desire reasoning) and all the social inference systems that create the logic of human social interaction—ones that enable as well as use inferences about the content of mental states—a broader conception of theory of mind may emerge: a computational model embodying a Theory of Human Nature (TOHN).

A fierce debate over the nature of the human mind has raged over the last two decades, and the study of reasoning has been a principal battleground. Broadly construed, reasoning is the ability to generate new representations of the world—new knowledge—from given or observed information. It is often considered constitutive of human intelligence: the most distinctly human cognitive ability, often thought to exist in opposition to, and as a replacement for, instinct. Discovering the nature of the inferential procedures whereby new knowledge is generated is, therefore, a foundational task of the cognitive sciences, with implications for every branch of the social sciences (Tooby & Cosmides, 1992).

One side of the reasoning debate has defended a long-standing and traditional view of the evolved architecture of the human mind: that it is a blank slate, a neurocomputational system equipped with content-free inferential procedures that operate uniformly on information drawn from all domains of human activity. This view implies that "nothing is in the intellect that was not first in the senses", as Aquinas famously put it—that is, all the mind's content originates in the world, and is built by content-free inferential procedures acting on data delivered by perceptual systems. Sometimes those procedures are thought to embody rational algorithms such as Bayes' theorem (Luce, 2003; Staddon, 1988; see Gigerenzer & Murray, 1987), multiple regression (Rumelhart & McClelland, 1986), or the inferential rules of propositional logic (Bonatti, 1994; Rips, 1994; Wason & Johnson-Laird, 1972): powerful algorithms thought to be capable of solving all problems and, therefore, not specialized for making inferences about any particular domain. In other cases reasoning is thought to be accomplished by heuristic procedures or rules of thumb, their complexity constrained by the size of working memory or the nature of perceptual representations (Kahneman, Slovic & Tversky, 1982; Kahneman, 2003). But whether these procedures are seen as powerful or error-prone, they are viewed as empty of content and, therefore, useful generally, no matter what subject matter (domain) one is called on to reason about. Hence they are known as *domain-general* reasoning procedures.

There seems little doubt that the evolved architecture of the mind contains some inferential systems that are (relatively) content-free and domain-general, although even in these cases the inferential procedures involved appear specialized for solving particular adaptive problems (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Brase, Cosmides & Tooby, 1998; Cosmides & Tooby, 1996, 2000; Gigerenzer, Todd, et al., 1999; Gallistel & Gibbon, 2000). The question was never whether *some* inferential procedures are relatively content-free and domain-

general, but whether all of them are—a claim that is central to the Standard Social Science Model (Tooby & Cosmides, 1992). The deep question about human nature is whether the evolved architecture of the mind *also* contains inferential procedures that are content-rich and domain-specific, ones that make inferences that go far beyond the information available to perception and not derivable from logic and mathematics alone.

Challenges to the view that most inferential procedures are content-free and domain-general began in the early 1980s, as cognitive developmentalists and evolutionary psychologists began studying the development and architecture of reasoning within particular content domains: reasoning about objects and their interactions (intuitive physics), about animals and plants (intuitive biology), about mental states (theory of mind), and about social interactions (e.g., social exchange). Researchers in these areas began to find evidence of inferential systems that looked like domain-specific natural competences. Their architecture and development seemed to shatter the age-old distinction between reasoning and instinct. These computational systems were equipped with proprietary, content-rich concepts/ representations and domain-specialized inference procedures, which reliably developed in the human mind in the absence of explicit instruction. Moreover, their representations and procedures were functionally specialized for solving a recurrent adaptive problem, and were dissociable from other, more general forms of reasoning. They were *reasoning instincts*. The inferences they made went far beyond what could be validly concluded on the basis of sense data alone—yet only when operating within a content-limited domain.

For example, it was shown that as early as infants could be tested—about two-months of age—they have preconceptions about what counts as an object (Baillargeon, 1987; Spelke, 1990) and know that two solid objects cannot pass through one another; by (at least) seven months, they make sophisticated causal inferences about object mechanics and launching events (Leslie, 1994). People everywhere organize the plant and animal world into a taxonomic hierarchy (Atran, 1990), and three year olds make accurate inferences about predatory-prey interactions whether they are raised in predator-impoverished Berlin or among jaguars and game animals in the Amazon (Barrett, 2005).

There seem to be domain-specific systems specialized for reasoning about the social world as well, represented by two different bodies of research: one on cognitive adaptations for social exchange (Cosmides 1985, 1989; Cosmides & Tooby 1992, 2005), the other on theory of

mind (Baron-Cohen, Leslie & Frith, 1985; Leslie, 1987; Baron-Cohen, 1995, 2005; Saxe, Carey, & Kanwisher, 2004). The former claims that our ability to reason about social exchange is generated by *social contract algorithms*: a set of programs that were specialized by selection to solve the intricate computational problems inherent in adaptively engaging in social exchange behavior, including cheater detection. The latter claims that we reliably develop a neurocomputational system designed for inferring that other people's behavior is caused by invisible entities that cannot be seen, heard, touched, smelled or tasted: *mental states*, including beliefs, desires, goals, and percepts (Leslie, 1987; Baron-Cohen, 1995). In both cases, claims for a reliably-developing domain-specialized inference system are based on evidence of (i) content-triggered functional dissociations, which reveal a design well-engineered for solving a specific adaptive problem; (ii) neural dissociations linked to brain damage and developmental disorders or revealed in brain imaging studies; (iii) robust precocious development; and (iv) cross-cultural uniformity (for reviews, see Cosmides & Tooby, 2005; Baron-Cohen, 1995, 2005; Saxe, Carey, & Kanwisher, 2004).

Both claims for a domain-specialized social inference system have of course been challenged. Evidence about neural correlates of theory of mind—the topic of this special issue—will surely contribute to the debate over whether inferences about mental states are caused by a system specialized for that function. Likewise, the study we report herein contributes to the debate about the domain-specificity of social exchange reasoning, testing a series of more domain-general alternative hypotheses: that social exchange reasoning is caused by a system for reasoning about all conditional rules (Almor & Sloman, 1996; Rips, 1994; Kirby, 1994, Oaksford & Chater, 1994; Johnson-Laird & Byrne, 1991), all familiar conditional rules (Goel, Shuren, Sheesley, & Grafman, 2004), all deontic conditional rules (i.e., rules involving obligation or entitlement; Cheng & Holyoak, 1985, 1989; Fodor, 2000; Sperber, Cara, & Girotto, 1995), or all deontic conditional rules involving subjective utilities (Manktelow & Over, 1991). When all the rules tested are familiar, these alternative hypotheses form a nested hierarchy of class inclusion, with the deontic+utilities hypothesis being the most domain-specific of the domain-general alternatives (see Figure 1). If that alternative fails, they all fail. So the question is, are there content-triggered neural dissociations *within* the class of deontic rules involving subjective utilities? In particular, do deontic rules involving social exchange engage different brain areas than deontic rules from other adaptive domains?

(Figure 1 about here)

Accordingly, the brain imaging study we conducted contrasts reasoning about familiar conditional rules drawn from three different content domains: (i) deontic rules about social exchange, (ii) deontic rules specifying what precautions ought to be taken in hazardous situations, and (iii) indicative rules describing people's preferences, habits, or traits (see Figure 2). All the conditional rules involved people's behavior and employed familiar content drawn from everyday life. Importantly, all reasoning problems were presented as Wason selection tasks, which ask subjects to identify which individuals may have violated a conditional rule. Using this format, we constructed reasoning tasks that place identical task demands on any auxiliary system activated while solving a word problem (working memory, vision, reading, etc). Therefore, any differences in brain activations should be attributable only to the content of the rules about which subjects are reasoning. If reasoning about social exchange is caused by a functionally isolable system, then it would be reasonable to expect different patterns of brain activation when reasoning about social exchange than when reasoning about other social rules, deontic or otherwise.

(Figure 2 about here)

The second purpose of the study we report is to determine whether, and at what stage, social exchange reasoning engages the theory of mind inference system. Research on neural correlates of theory of mind suggests that making inferences about mental states engages medial prefrontal cortex, anterior temporal cortex including the poles, and posterior temporal cortex including the superior temporal gyrus and temporo-parietal junction (e.g., Apperly, Samson, Chiavarino, & Humphreys, 2004; Baron-Cohen et al., 1999; Fletcher et al., 1995; Gallagher et al., 2000; German, Niehaus, Roarty, Giesbrecht, & Miller, 2004; Saxe & Kanwisher 2003; Saxe & Wexler, 2005; for reviews, see Frith & Frith, 2003; Gallagher & Frith 2003; Saxe, Carey, & Kanwisher, 2004). Less consistently implicated areas include orbitofrontal cortex (Stone, Baron-Cohen, & Knight, 1998), amygdala (Baron-Cohen et al., 1999; Fine, Lumsden, & Blair, 2001; Stone, Baron-Cohen, Calder, Keane, & Young, 2003), and posterior cingulate (Fletcher et al., 1995). The design of our study allows one to see whether these neural correlates of theory of mind are most activated when subjects are recognizing/ interpreting rules as involving social exchange, or during the post-interpretive stage where potential cheaters are being identified. Three previous brain imaging studies have investigated reasoning about social exchange

(Canessa et al., 2005; Fiddick, Spampinato, & Grafman, 2005; Wegener et al., 2004), but this is the only one designed such that neural activations during these different stages of processing can be distinguished.

Theory of mind and strategic social interaction. The adaptive function of the theory of mind system is often described as (i) predicting and explaining behavior in terms of mental states, and (ii) inferring their content on the basis of cues (e.g., eye direction for inferring desire or object of attention; typical actions with wrong object for pretense; failed goal-directed action for intended goal or false belief content). But the discussion often stops at that point, leaving one with the impression that the machinery performing these computations was selected for because it furthered the pure beauty of contemplation. That cannot be true, of course: Natural selection does not build complex functional computational systems unless they contributed to adaptive behavior in some way. So the question is, how did inferring the content of other people's mental states contribute to adaptive behavior in ancestral environments?

Baron-Cohen (1995) argues that strategic social interaction—situations in which the best behavioral strategy for me to pursue depends on what you intend to do—creates selection pressures favoring computational machinery for inferring the content of other people's mental states (also Humphreys, 1976). Evolutionary biologists have used game theory to analyze strategic social interaction, with the goal of discovering which decision rules are likely to have evolved—that is, which implement an evolutionarily stable strategy (ESS). An ESS is a strategy (a decision rule) that can arise and persist in a population because it produces fitness outcomes greater than or equal to alternative strategies (Maynard Smith, 1982). ESS analyses have illuminated strategic social interaction in many domains, including parental care, mating, dominance interactions, threat, communication, foraging, collective action, and social exchange (for review, see Maynard Smith, 1982; Gintis, 2000).

In social exchange, individuals agree, either explicitly or implicitly, to abide by a *social contract*, a situation in which an agent is *obligated* to satisfy a requirement of some kind (often at some cost to the agent), in order to be *entitled* to receive a *benefit* from another agent. These understandings can be expressed as conditional (*If-then*) rules that fit the following template: *If you accept a benefit from agent X, then you are obligated to satisfy X's requirement*. In social exchange, a *cheater* is an individual who intentionally violates a social contract by taking the benefit specified without satisfying the requirement that provision of that benefit was made

contingent on. Modeling selection pressures for social exchange as a repeated prisoner's dilemma (Trivers, 1971; Axelrod & Hamilton, 1981; Boyd, 1988), evolutionary biologists have shown that rules of reasoning and decision making that guide social exchange in humans will implement an ESS only if they include design features that solve an intricate series of computational problems (Cosmides & Tooby, 1989). Of these, the ability to detect cheaters has received the most attention (Cosmides, 1985, 1989; Cosmides & Tooby, 2006). But the ability to infer the contents of other individual's mental states—especially what they *want* and what they *intend* to do—is also essential for social exchange to evolve (Cosmides, 1985; Cosmides & Tooby, 1989). Social exchange depends on the ability to infer the content of other people's *desires*, *goals*, and *intentions*.

For example, let's say that my *goal* is to go downtown so I *want* to borrow your car. Knowing that I want to borrow your car—that I consider it a benefit—you could agree to lend it to me, but only on the condition that I fill the tank. That is, you could offer the following social exchange: "If you borrow my car, then you must fill the tank with gas". But you could not offer me this exchange unless you had inferred the content of one of my mental states, forming the representation *agent-wants*-[to borrow car], from what I say, from my eye direction, from my behavior, or from my having no means to achieve my *goal*. Without the ability to infer the content of *your* goals or desires (viz. that you consider having a full tank of gas a benefit; i.e., desirable) I would not recognize that your suggestion fits the social contract template; after all, "If you borrow my car, I'll break your legs" is a threat, not an offer to exchange, and is recognizable as such because breaking my legs is a cost to you as well as to me. Accordingly, reasoning research using the Wason selection task shows that cheater detection is triggered when a deontic rule fits the benefit-requirement template of a social contract, but performance suffers when the action to be taken is not a benefit (Cosmides & Tooby, 1992, 2005; Cosmides, Barrett & Tooby, forthcoming).

Without recognizing that an offer to satisfy my desire conditional on my satisfying yours fits the social contract template, I would not be able to make correct inferences from your offer—for example, that it also implies that if I fill your tank then I am *entitled* to borrow your car. This inference is licensed by a domain-specialized grammar of social exchange that operates on content-rich representations of benefits and requirements of agents, and subjects spontaneously make it (Cosmides, 1989; Fiddick, Cosmides & Tooby, 2000; Gigerenzer & Hug,

1992). But it is not licensed by the benefit-less, agent-less, content-free rules of propositional logic. Consistent with the claim that social contract reasoning dissociates from logical reasoning, Maljkovic (1987) found that individuals with schizophrenia had deficits in logical reasoning (ones consistent with frontal lobe impairment), yet reasoned normally when asked to detect cheaters on Wason tasks involving social exchange.

The ability to infer intentions is also necessary. For example, my agreeing to fill your tank depends on my infering that you do not *intend* to lend me your car otherwise. More significantly, social exchange is difficult to evolve without the ability to distinguish intentional cheating from noncompliance due to accidents or innocent mistakes (Panchanathan & Boyd, 2003). To implement an ESS, social contract algorithms must be good at detecting individuals equipped with *designs* that cheat, so those individuals can be excluded from future interactions. But a strategy that refuses to cooperate with individuals who violated a past social contract by accident loses many opportunities to gain from cooperation; simulation results show such strategies get selected out in the presence of strategies that exclude only intentional cheaters. Accordingly, reasoning research shows that cheater detection is triggered only by intentional violations of social contracts, not by innocent mistakes (Fiddick, 2004; Cosmides, Barrett & Tooby, forthcoming; Cosmides & Tooby, 2005).

The grammar of social exchange can, in turn, support inferences about the content of mental states. Third parties—including subjects in reasoning experiments—could not recognize your offer as a social contract unless they knew the contents of both of our desires: that I want to borrow your car, and that you want a full tank of gas. Notice, however, that what *you* want can be inferred from the fact that you are offering a conditional benefit to *me*—the structure of interaction in social exchange implies that what you require in exchange for providing a benefit to me is something you *want*: help, goods, or a state of affairs. The logic of social exchange allows the contents of mental states to be inferred, and inferring mental states allows social exchange to proceed. This implies that we can think of theory of mind in a broad, rather than a narrow, sense. TOM(narrow) refers to a small range of inferences: using beliefs and desires to predict and explain behavior, inferring knowing from seeing, inferring wanting from eye direction. But TOM(broad)—or perhaps TOHN, a Theory of Human Nature— would include *all* social inference systems that create the logic of human social interaction, ones that enable, as well as use, inferences about the content of mental states. From this point of view, studying

neural correlates of social exchange reasoning *is* studying neural correlates of theory of mind—of Theory of Mind(broad).

Reasoning about precautionary rules. As the prior analysis shows, reasoning about social exchange is not possible without mental state inferences. The same is not true of reasoning about precautionary rules, which fit the template *If you engage in hazardous activity H, then you must take precaution P* (e.g., "If you work with TB patients, then you must wear a surgical mask"). Precautionary rules are deontic because they specify what you *ought* to do in a given situation. According to Fiddick et al. (2000), the function of detecting violations of precautionary rules is to manage risk—to tell when someone or something is in danger by virtue of having not taken appropriate precautions (see also Boyer & Lienard, in press). Inferences about intentionality are therefore unnecessary: you are in danger from having violated a precautionary rule, whether you violated it on purpose or by accident. Accordingly, Wason tasks involving precautionary rules elicit high levels of violation detection, whether the violation is intentional or accidental (Fiddick, 2004). This result is in contrast to social exchange, where subjects are good at detecting intentional violations but not accidental ones.

Recognizing and interpreting precautionary rules does not require inferences about the content of anyone's goals or desires—you can find yourself in a hazardous situation whether you want to be there or not. Interestingly, R.M., a patient with some deficits on theory of mind tasks (he had bilateral damage to medial orbitofrontal cortex and anterior temporal cortex, including disconnection of both amygdalae) was very good at detecting violations of precautionary rules in Wason selection tasks. Yet his ability to detect violations of social contracts in the Wason task was severely impaired (Stone, Cosmides, Tooby, Kroll & Knight, 2002). The two sets of tasks were logically isomorphic with identical task demands. Normal subjects performed equally on both, yet were not at ceiling, ruling out the possibility that ceiling effects were masking real differences in difficulty. Under these circumstances, a single dissociation is evidence that social exchange and precautionary rules are activating somewhat different brain systems.

Precautionary rules versus social contracts: how many domain-specific mechanisms? In social exchange, benefits are delivered conditionally; it therefore requires conditional reasoning for its regulation. The Wason selection task is a standard tool for investigating conditional reasoning. Subjects are given a rule of the form *If P then Q*, and asked to identify possible violations of it—a format that easily allows one to see how performance varies as a

function of the rule's content. Performance is usually poor when the rule is indicative, describing some aspect of the world: only 5-30% of normal subjects choose the logically correct answer, *P & not-Q*, for these descriptive rules, even when they relate familiar content drawn from everyday life (Manktelow & Evans, 1979; Cosmides, 1985, 1989; Wason, 1983). In contrast, 65-80% of subjects answer correctly when the rule is a social contract and a violation represents cheating. The same is true for precautionary rules (Cheng & Holyoak, 1989; Fiddick et al., 2000; Manktelow & Over, 1988, 1990; Stone et al., 2002).

Not all deontic rules elicit high levels of performance (Cosmides, 1989; Cosmides & Tooby, 1992, 2005; Cosmides, Barrett & Tooby, forthcoming). But the pattern elicited by deontic rules that are social contracts and precautions is so different from that elicited by indicative ones that most theories in the reasoning literature have features designed to explain a deontic-indicative difference—but not a social contract-precautionary difference. Judging precaution violations and detecting cheaters on a social contract are so alike that, according to alternative theories, the cognitive architecture of the human mind does not distinguish between them (e.g., Buller, 2005; Cheng & Holyoak, 1985, 1989; Cummins, 1996; Fodor, 2000; Kirby, 1994; Oaksford & Chater, 1994; Johnson-Laird & Byrne, 1991; Manketlow & Over, 1991; Rips, 1994; Sperber, Cara, & Girotto, 1995). Like social contracts, precaution rules are conditional, deontic (they express the conditions under which a person is permitted to take action X, or ought to take precaution Y), and involve subjective utilities (i.e., perceived benefits and costs).

An alternative view is that the mind contains a functionally distinct, domain-specific cognitive specialization for reasoning about hazards, as well as a social contract specialization (Cosmides & Tooby, 1997; Fiddick, 2004; Fiddick et al., 2000; Stone et al., 2002). Given the evidence of distinct reasoning mechanisms from the behavioral and patient data, we sought to explore the neural correlates of normal subjects' reasoning about social exchange, precautionary rules, and familiar, indicative rules describing social behavior. Recent fMRI studies have shown dissociation of social contract reasoning from precautionary reasoning (Fiddick et al, 2005; Wegener et al., 2004) and from general descriptive rule reasoning (Canessa et al, 2005). Our study included all three types of reasoning in the same experimental paradigm, using the Wason selection task. This task may be ideal for imaging studies in that the three types of reasoning problems differ only in their content. Not only are the task demands identical across problem type, but the performance of normal subjects on the social contract and precautionary rules tested

is identical, both in percent correct and reaction time (i.e., these problem sets do not differ in difficulty). We had subjects read stories describing a social exchange, precautionary, or descriptive (indicative) social rule and respond "yes" or "no" to whether various instances were possible violations of that rule. Both the behavioral and patient data suggest that social exchange and precautionary reasoning should show different patterns of brain activations, and that both of these types of reasoning should activate different areas than reasoning about indicative rules, even when these describe social behavior.

We examined brain activations in response to both reading the stories (where interpretation of the rules plausibly happens) and determining possible rule violations (information search leading to detection of violations). There were two reasons for this. First, some theories strongly distinguish the interpretive process from post-interpretive information search; for example, Sperber et al. (1995) view the post-interpretive process as reflecting nothing more than a domain-general ability to categorize (proposals by Fodor (2000) and Buller (2005) are similar in this regard). Second, we thought social contract algorithms would be more likely to engage the theory of mind system during the interpretive process, for the reasons discussed above. Although cheater detection is only activated by the possibility of intentional violations, information about intentionality was presented in the stories, not on the cards.

Brain areas of interest. R.M., the patient with a selective deficit in social exchange reasoning, had suffered bilateral damage to medial orbitofrontal cortex and anterior temporal cortex; damage to his anterior temporal poles and perirhinal cortex was so severe that both amygdaloid complexes were disconnected. All three areas have been implicated in theory of mind reasoning, and hence are areas of interest.

Using a very different design that varied modal operators and order of antecedent and consequent in the conditional, Fiddick et al. (2005) report greater activation for social contracts than precautions in the dorsomedial prefrontal cortex (BA 6,8), Wegener et al. (2004) report social contracts elicit greater activation relative to precautions in bilateral anterior prefrontal cortex (PFC; BA 10, 11), dorsomedial PFC (BA 6, 8), left posterior temporal cortex (BA 22), and left parietal cortex (BA 40). In comparing unfamiliar social contracts to social descriptive rules, Canessa et al. (2005) report social contracts elicit greater activation in dorsomedial PFC (BA 8), left anterior (BA 46) and right posterior (BA 9) PFC, and right parietal cortex (BA 39). None of these designs allow one to distinguish interpretation from violation detection.

Based on neuroimaging during syllogistic reasoning tasks, Goel, Dolan, and colleagues (Goel, Shuren, Sheesley, & Grafman, 2004; Goel, Buchel, Frith, & Dolan, 2000; Goel & Dolan 2001; 2003; 2004) suggest that there are two distinct systems for reasoning: one employed for abstract or unfamiliar content (bilateral fronto-parietal system), and one employed for familiar content (left lateral fronto-temporal system). In a neuropsychological study, Goel et al. (2004) suggest that these brain networks should extend to other modes of deductive reasoning, specifically the Wason selection task. However, it has been known for some time that familiar social content is neither necessary nor sufficient to elicit good reasoning performance on the Wason selection task (Cosmides, 1989; Manktelow & Evans, 1979; Wason, 1983). It turns out that the "familiar" Wason task used by Goel et al. (2004) was a social contract, and they found performance was particularly impaired in patients with damage to part of the fronto-temporal system, the left frontal lobe.

Method

Subjects

Twenty healthy graduate students and staff members at Dartmouth College were paid $20 for their participation. Volunteers were screened for medication use, history or neurological or psychiatric disorders, and other serious medical conditions. Because we were interested in brain areas involved in successful reasoning about social contract and precautionary rules, four subjects were excluded from the analysis for poor behavioral performance on the task (<50% correct on all three types of problems). Four additional subjects were also excluded from the analysis: two for lost data due to technical malfunctions, one for excessive head movement, and one due to structural abnormalities that caused problems during spatial normalization. These exclusions left 12 subjects (4 males, mean age = 24.6 years, SD = 3.29) in the analysis. This research was approved by the UCSB Human Subjects Committee and the Dartmouth Committee for the Protection of Human Subjects, and all subjects gave informed consent.

Materials

Eight social contract, eight precautionary, and eight descriptive rule Wason selection tasks were used in this study[1] (see examples in Figure 3 and Appendix). All the conditional rules involved people's behavior and employed familiar content drawn from everyday life. Problems were very closely matched on word length (social contract: M = 166.8, SD = 15.1, Range = 147-183; precaution: M = 166.3, SD = 11.1; Range = 152-182; descriptive: M = 166.6, SD = 11.3;

Range = 152-184). Before use in imaging, they were normed on 56 undergraduates at the University of California, Santa Barbara. The social contract and precautionary problems were matched on performance in undergraduates (81.7% and 83.5% correct, respectively, N = 56; correct = choosing *P*, *not-Q*, and no other cards). Furthermore, all descriptive rules were about people—their habits and behavior—but they did not fit the functional logic of either social contract or precaution rules. As is typical, performance by undergraduates on descriptive rules (M = 42.8% correct) was lower than for social contracts or precautions.

Task

The Wason selection task was composed of two parts: the story that presents the rule and the cards that ask subjects about potential violations. This design was employed to be able to examine separately the brain areas involved in rule interpretation (reading the story) and decision-making (responding to the cards).

The stories were presented in three parts (see Figure 3). The first part (panel A of Figure 3) introduced and gave a rationale for the rule, and specified whether the concern was about people cheating on the rule (social contract), breaking the safety rule (precaution), or simply violating the rule (descriptive). The second part (panel B) explained the cards, specified what information was contained on them, and reiterated the concern: may have cheated (social contract), may be in danger (precaution), or rule may be wrong (descriptive). (Specifying the concern was to trigger the intended domain.) The third part (panel C) presented the rule again and explained that the task was to find out which people had violated the rule (with no mention of rule type or of concerns with cheating, danger, or wrongness). These parts of the story were presented for 15.0 sec, 12.5 sec, and 7.5 sec, respectively. Each story was preceded by a 2.5 sec prompt indicating that the story was about to appear. Thus, the story presentation lasted 37.5 sec in total for each story.

Following the story, eight cards (two for each logical category: P, not-P, Q, not-Q) were presented individually, one at a time, with the question “Could this person have violated the rule?”. Each card gave information about what a particular individual did or did not do, with each individual mentioned only once (e.g., Jake did P; Maya did not do Q; see panels E and G in Figure 3 and Appendix). The rule was presented on the screen to avoid excessive demands on memory. Each card was presented for 5 seconds, during which subjects were asked to respond “yes” or “no”, using a right or left button press, to whether each instance was a possible violation

of the rule. Cards were presented in a random order. Card presentation was jittered such that the time between each card varied among 0, 2.5, or 5 seconds. The card presentation period lasted a total of 55 sec for each problem, including the jitter time.

Before being scanned, subjects completed a practice set of three problems on a laptop computer. This method was followed to ensure that subjects understood the instructions and would be familiar with the format of the problems and the required responses. Subjects were given unlimited time to complete the first two practice problems. The third practice problem was displayed for the times used in the scanner (as shown in Figure 3).

(Figure 3 about here)

Imaging

Subjects completed four functional runs, each consisting of six Wason selection task problems (two each of social contract, precaution, and descriptive), and two rest (fixation cross; 20 sec each) periods. Order of runs was counterbalanced across subjects. Order of problems and rest periods within each run and order of cards within each problem were randomized for each subject. Each functional run ended with 20 sec of fixation cross, and lasted a total of 10.25 minutes.

Anatomical and functional images were acquired with a 1.5-T whole body scanner (General Electric Medical Systems, Signa, Milwaukee, WI) with a standard head coil. Foam pads were used to minimize head movement. Stimuli were presented using a laptop running PsyScope (Cohen et al., 1993). Subjects viewed stimuli projected onto a screen through a mirror mounted on the head coil. Responses were made using two magnet-compatible fiberoptic button presses, one per hand, which interfaced with the PsyScope Button Box (Carnegie Mellon University, Pittsburgh, PA). Anatomical images were acquired using a high-resolution 3-D spoiled gradient recover sequence (T1, TE = 6 msec, TR = 2500 msec, flip angle = 25°, 124 sagittal slices, voxel size = 1 x 1 x 1.2 mm). Functional images were acquired using a gradient spin-echo, echo-planar sequence sensitive to BOLD contrast (T2*, TR = 2500 msec, TE = 35 msec, flip angle = 90°, 3.75 mm x 3.75 mm in-plane resolution), using 25 interleaved 4.5 mm axial slices (1-mm skip between slices) to image the whole brain. Each subject was scanned for four functional runs of 246 repetitions. The first six functional images from each functional run were dropped to allow the signal to stabilize.

Analysis

Images were preprocessed using SPM2 (Wellcome Department of Imaging Neuroscience, University College London, UK). We registered all functional images to the first volume to correct for minor head movements and then to the anatomical image. Images were transformed to the MNI brain template, and functional images were spatially smoothed using an 8 mm FWHM Gaussian filter.

Subsequent analysis was conducted using custom software written in MATLAB (The MathWorks, Natick, MA). The general linear model was used to analyze the fMRI time-series (Friston et al., 1995). Our methods for modeling the response to cards followed those of Ollinger et al. (2001). For cards, each stimulus onset and post stimulus time point (up to a specified limit, in this case 17.5 sec.) was modeled by a separate parameter. There were seven post-stimulus time bins covering a total window length of 17.5 seconds. This approach is very similar to selective averaging (Dale & Buckner, 1997) in that it can be thought of as selective averaging without counterbalancing of trial orders. This model is also known as a finite impulse response model (Henson et al., 2001). The benefit of this model is that it makes minimal assumptions about the shape of the hemodynamic response, thus accommodating variations in the timing of the response that have been observed across brain regions (e.g., Schacter et al., 1997) and avoiding the amplitude bias that these variations can introduce (Calhoun et al., 2004). A related method was used for modeling the response to stories. We assume that the response to a story reaches a stasis at the 7th post-stimulus time point at the latest. Accordingly, six consecutive time bins modeled the rise of the response, a single "box car" modeled the stasis of the response lasting until the offset of the story, and six additional time bins modeled the fall of the response. As before, this approach makes minimal assumptions about the shape of the hemodynamic response. The response to cards and stories was modeled separately for each of the three content types (descriptive, precautionary, and social contract), producing a total of six estimated responses. In addition to the parameters already discussed, four parameters modeled linear drift within each session and four parameters modeled the session-specific means.

A group level random effects model was conducted. For the purposes of contrasting the response to different card types, activation levels for each of the three types (descriptive, precautionary, and social contracts) were estimated by summing the estimated hemodynamic response along the interval of 2.5 to 15 seconds post stimulus onset. These sums were then submitted to the group level analysis. For contrasts between story types, the box car parameter

modeling the sustained response to the story was submitted to the group level analysis. We compared each of the three types of problems (social contracts, precautions, and descriptives) to one another separately for the Wason stories and for the cards. We also compared the average of all three types to baseline (fixation cross) for both stories and cards. All contrasts were bidirectional, using a threshold of $p < .005$ (2 tailed; uncorrected). To control for false positives, activations were not considered significant unless a cluster of 10 contiguous voxels survived the threshold.

To further address the false positive issue, we examined the average signal intensity for each individual problem for the major activations we found in the story comparisons. Activation levels were obtained by (1) averaging the signal across all voxels contained in the cluster defined by the group comparison; (2) removing linear drift and session-specific effects; (3) averaging the signal between 15 and 35 sec post stimulus onset individually for each story; (4) averaging these results across subjects. If the differential activation does in fact reflect differences in the underlying representations of these problems, then it should replicate *across individual problems*. That is, we should see a consistent separation between problem types, with (e.g.) most social contracts activating a particular area more than most precautions. Evidence of such a separation yields more confidence that the activation difference is real, rather than an artifact of a few problems.

The problem of false positives increases when contrasting conditions place different processing demands (e.g., recognition versus recall memory). In contrast to many fMRI studies, the contrasts here are for conditions that place identical task demands (they are all Wason tasks) and the behavioral data for social contracts and precautions are indistinguishable. That the conditions are so closely matched is an added, theoretical control on false positives.

Results and Discussion

Behavioral Results

Were social contract and precaution problems well-matched? Yes: planned contrasts showed that subjects performed equally well on social contract (90.6% correct, SD = 12.1) and precaution rules (91.7% correct, SD = 14.4; $F(1,11) = 0.61$, $p = .81$). Thus, our performance criterion for inclusion in the study (>50% correct on social contracts and precautions) ensured that overall performance was quite high (cf. to undergraduate performance of 81.7% and 83.5% correct for social contracts and precautions, respectively, N = 56). Performance on descriptive

rules (59.4% correct, SD = 34.6) was significantly worse than performance on social contracts and precautions ($F(1,11) = 10.76$, $p = .007$), but was still better than undergraduate performance on these same problems (42.8% correct, N = 56). Likewise, mean reaction time to social contracts (1883 msec, SE = 100.8) and precaution rules (1880 msec, SE = 96.7) was identical ($F(1,11) = 0.01$, $p = .94$), whereas mean reaction time to descriptive rules (2111 msec, SE = 126.0) was slower than for social contract and precaution rules ($F(1,11) = 12.50$, $p = .005$).

Imaging Results

Manipulation Check. As a check on our methods, we compared activations across stories (collapsed across content) and across cards (collapsed across content) to rest (fixation cross). As expected for cognitive tasks requiring attentive processing (for review, see Cabeza & Nyberg, 2001), stories strongly activated bilateral visual cortex, left temporo-parietal and left posterior parietal regions compared to rest (stories > rest; see Table 1). Smaller activations were seen in left dorso-lateral prefrontal cortex. Also as expected, there were typical deactivations relative to rest in medial prefrontal and cingulate areas (rest > stories; see Table 1). Activation patterns for card choice (collapsed across content) compared to rest (fixation cross) produced strong bilateral activations in dorso-lateral prefrontal cortex, superior posterior parietal cortex, and visual cortex. Smaller activations were seen in bilateral temporo-parietal regions (cards > rest; see Table 2). Deactivations relative to rest were seen in medial prefrontal cortex and anterior and posterior cingulate cortex (rest > cards; see Table 2). Thus, stories produced greater activations in left temporo-parietal regions and cards produced greater activations in bilateral dorso-lateral prefrontal cortex, supporting a distinction between interpretive processing and decision making.

(Table 1 about here)

(Table 2 about here)

Story contrasts. Figure 4 illustrates activations for contrasts among all three story types. Story contrasts produced strong activations and robust time courses (see examples in Figure 5). Reading social contract stories, relative to precaution stories, activated right anterior temporal (BA 20) and left posterior temporal (BA 21) cortex, a lateral prefrontal area on the right (BA 6), and posterior cingulate (BA 23; see Table 3 and Figure 5). Compared to descriptive stories, social contracts activated anterior temporal cortex bilaterally (BA 22). Reading precaution stories produced greater activations than social contracts in left dorso-frontal (BA 6, 9) and parietal (BA 2) regions, and areas of the cingulate (BA 31, 32; see Table 4). Compared to descriptives,

precaution stories activated ventro-lateral prefrontal (BA 10) regions bilaterally, superior parietal (BA 7, 40) regions on the right, and posterior cingulate (BA 31). Descriptive stories, relative to precautions, activated the right parahippocampal gyrus at the amygdala (see Table 5). The descriptive minus social contract contrast produced no significant clusters.

(Figure 4 about here)

Card choice contrasts. Overall, the card contrasts produced weaker activations and less robust time courses than the story contrasts did. This difference may result from the short duration of the card events compared to the much longer duration of the stories. No clusters survived threshold for the social contract minus precaution and the social contract minus descriptive card contrasts. However, responses to precaution cards compared to social contract cards showed activations in middle and ventral prefrontal (BA 6, 9, 46, 47), middle and posterior temporal (BA 21, 41), and superior occipital (BA 18, 37) regions, as well as the right insula (BA 13) and cingulate (BA 24; see Table 4). In contrast, compared to descriptive cards, responses to precaution cards showed greater activation only in dorso-lateral frontal regions: the precentral gyrus bilaterally (BA 4, 6) and the right postcentral gyrus (BA 6). Responses to descriptive cards showed greater activation compared to social contract cards in dorso-medial (BA 8) and dorso-lateral (BA 9) prefrontal areas, as well as in the cingulate gyrus (BA 24, 31), and right fusiform gyrus (BA 18; see Table 5). Compared to precaution cards, descriptives showed greater activation in dorso-lateral prefrontal cortex (BA 9, 46) and left parietal (BA 7) regions.

The interpretive process: Do social contract stories activate different areas than precaution stories? Yes, supporting the claim that social contracts and precautionary rules are interpreted via two different, functionally distinct, domain-specific inferential systems.

According to most theories, the same inferential processes interpret all deontic rules, whether they are social contracts, precautions, or some other species of permission rule. These inferential processes would be activated while subjects are reading the stories, and would result in social contracts and precautionary rules being given the same interpretation. If this were a correct description of what is happening, then the same brain areas should be activated whether subjects are reading social contract or precautionary stories, resulting in no differential activations for either the social contract > precaution or the precaution > social contract comparisons. Yet these comparisons did reveal differential activations. When subjects were interpreting the social contract stories, several areas commonly implicated in theory of mind

tasks were activated: the right anterior temporal cortex (BA 20), left posterior temporal cortex (BA 21), and the posterior cingulate (BA 23; see Figure 5 and Table 3).

(Figure 5 about here)

Figure 6A shows the average signal intensity in the anterior temporal cortex for each individual social contract and precautionary problem: Importantly, there is almost no overlap for these two sets of rules (Mann-Whitney $U = 9$, $p = .005$). Yet the surface content of the social contracts (the specific actions or items mentioned) varies widely; the social contracts are similar to one another only by virtue of fitting the benefit-requirement template shown in Figure 2A (correspondingly for precautionary rules, see Figure 2B). That the pattern of differential activation replicates across individual problems increases our confidence that what we are seeing is not an artifact of a few problems, but instead reflects the underlying, content-specific representation of social exchange versus precautionary problems. This anterior temporal activation for social contracts is consistent with the neuropsychological data from patient R.M., who was selectively impaired on social contract reasoning relative to precautionary reasoning (Stone et al., 2002). Panels B (middle temporal cortex; Mann-Whitney $U = 4$, $p = .001$) and C (posterior cingulate; Mann-Whitney $U = 15$, $p = .041$[2]) show a similar pattern of replication across individual problems.

(Figure 6 about here)

When subjects were interpreting precautionary stories, areas of dorso-medial prefrontal cortex (BA 6, 9) and the right cingulate gyrus (BA 31) were more active than when they were interpreting the social contract stories (see Figure 4 and Table 4). The analysis of individual precaution and social contract problems reveals that the two cingulate clusters show the cleanest separation between problem-types ($U = 5$, p = .001; U = 8, p = .005)

(Table 3 about here)

Another way of addressing the same question is to see what brain areas are activated by social contracts and precautions when each is compared to the exact same control condition: the descriptive rules. Compared to descriptives, social contracts activated anterior temporal cortex bilaterally (BA 22; see Figure 4 and Table 3). This activation is similar to what was found when social contracts were compared to precautions, and is likewise consistent with the neuropsychological data. The precaution > descriptive comparison also showed similarity with the precaution > social contract activations: the right cingulate gyrus (BA 31) was again active.

Ventro-medial prefrontal cortex (BA 10) was also activated for precautions relative to descriptives (see Figure 4 and Table 4). As Table 4 shows, interpreting precautionary rules did not activate areas typically associated with theory of mind, whether they are being compared to social contract or descriptive activations.

(Table 4 about here)

Taken together, these results indicate that different brain areas activate when subjects are interpreting social contract rules than when they are interpreting precautionary ones. Interpreting social contracts activates areas that have been associated with theory of mind—anterior temporal, posterior temporal, and posterior cingulate areas; anterior temporal activations were found for social contract stories compared to both precautionary and descriptive ones.

The decision-making process: Do different brain areas activate as a function of what type of violation one is looking for? The extreme view, advocated by Sperber (Sperber, Cara & Girotto, 1995; Sperber & Girotto, 2002), is that violation detection on the Wason task is mere categorization: during interpretation, the relevant values are computed—*P* and *not-Q* for deontic rules, *P* and *Q* for descriptives of the kind we have here—and then cards are categorized as to whether they match either of these values. This framework implies there will not be differential activations during the card choosing phase for deontic rules (i.e., social contracts and precautions will activate identically). Indeed, if matching to category is all that is at stake, it also implies no difference between deontic and descriptive rules during the card choice phase. In contrast, the domain-specific view implies that deontic rules are not all the same: looking for cheaters on a social contract engages different computational processes than looking for people who are in danger from having violated precautionary rules. It also implies that social contracts and precautions will activate differently than descriptives, even though all of these rules involve the behavior of people. Analyzing brain activations during the card decision phase can address these predictions.

Activations during the decision-making process were not the same for all deontic rules. When subjects were detecting violations of precautionary rules, a number of brain areas activated more strongly than when they were detecting violations of social contracts (see Table 4). These areas include portions of right dorso-lateral prefrontal (BA 46) and left ventro-lateral prefrontal (BA 47) cortex, the right insula (BA 13), medial cingulate (BA 24), and left middle (BA 21) and posterior temporal cortex (BA 41). No clusters survived threshold for the social contract >

precaution comparison, but this was also true for the social contract > descriptive one. This result should not be construed as indicating no difference between violation detection for descriptives and social contracts: compared to social contracts, detecting violations of descriptive rules more strongly activated several areas, including dorso-medial prefrontal cortex (BA 8, 9) and medial cingulate (BA 24, 31; see Table 5).

Compared to descriptives, precaution violation detection more strongly activated the areas along the precentral and postcentral gyri (BA 3, 4, 6; see Table 4). The descriptive > precaution comparison indicates activation of dorso-lateral prefrontal (BA 9, 46) and left superior partietal (BA 7) cortex (see Table 5).

These results suggest several things. First, the decision making process activates different areas, depending on whether the subject is looking for violations of precautionary rules, social contracts, or descriptives. Second, there is no evidence that the decision-making process activates theory of mind areas for social contracts—no more strongly, at least, than detecting violations of precautionary or social descriptive rules does. That ToM areas are activated during social contract interpretation but not during violation detection makes sense: computing other people's desires is necessary for recognizing that a conditional expresses a social contract. But once this mapping of agents' desires has occurred, cheater detection can proceed without these desires being re-computed.

Which brain areas are activated during reasoning about descriptive rules involving social behavior/ person-traits? During interpretation, social contracts activated a number of brain areas more strongly than descriptive rules did, but the reverse was not true: during interpretation, the descriptive > social contract comparison yielded no differences. This result could reflect the fact that both types of conditionals were about social behavior, but the social contracts required further social processing than the social descriptives. This interpretation is supported by the descriptive > precaution story comparison (see Table 5). During interpretation, descriptive stories activated an area implicated in social reasoning, the right parahippocampal gyrus at the amygdala.

(Table 5 about here)

As discussed above, during violation detection, descriptive rules activated different brain areas than both social contracts and precautions. We note that areas activated by violation

detection for descriptive rules include ones usually associated with more deliberative forms of reasoning (i.e., areas of dorso-lateral prefrontal cortex, Goel & Dolan, 2004; Goel et al., 2004).

Areas of overlap. Our claim that social exchange and precautionary reasoning produce different patterns of brain activation should not be construed as implying that *no* brain areas are activated by both. Indeed, the tasks were designed to be very closely matched on any dimension that could affect auxiliary systems such as working memory or attentional resources. Accordingly, the results discussed in the manipulation check imply that all the tasks activated areas involved with reading and decision making.

Concordance with other studies? Although the methods used across studies comparing social exchange to precautionary reasoning were very different, there were a few areas of concordance. Wegener et al. (2004) and Fiddick et al. (2005) both report social contracts differentially activating dorso-medial PFC within BA 6; social contracts activated a right lateral portion of BA 6 in our study. Wegener at al. (2004) report posterior temporal activation on the left in BA 22 for social contracts; we found a similar temporal activation in this area for the social contract > precaution comparison (BA 21), as well as anterior temporal areas (R BA 20 for social contract > precaution; bilateral BA 22 for social contract > descriptive comparison).

Conclusion

Managing hazards and engaging in social exchange pose very different adaptive problems—different enough that the computational requirements of a system well-engineered for making adaptive inferences about social exchange are incommensurate with those of a system well-engineered for reducing risks in hazardous situations (Cosmides & Tooby, 1989, 1997; Fiddick, 1998, 2004; Fiddick et al., 2000). For that reason, it had been proposed that two functionally distinct neurocomputational specializations evolved, one for reasoning about social exchange and the other for reasoning about precautionary rules. The neuroimaging results reported here add to the set of behavioral and neuropsychological dissociations supporting that hypothesis. Equivalent reasoning problems, matched on task demands and difficulty, elicited different patterns of brain activation depending on whether their content involved social exchange or taking precautions against hazards. This was true during the phase in which subjects were interpreting the rules, as well as during the post-interpretive phase in which they were deciding which individuals could have violated these rules.

In other words, the results revealed content-triggered neural dissociations *within* the already narrow class of deontic rules involving utilities. Different patterns of neural activation for social contracts and precautions should not exist if more domain-general theories of reasoning were correct. According to those theories, precautionary and social exchange rules are just instances of a more general class of conditional rules, such that both are operated on by the same neurocomputational machinery. These theories differ only in their claims about which more general class this machinery is designed to operate on (see Figure 1). For some, it is the class of all deontic rules: social contracts and precautions are said to be interpreted as fitting the template of a permission schema (Cheng & Holyoak, 1985; 1989) or assigned the same logical form using deontic operators, such as *forbid(P and not-Q)* (Sperber et al., 1995) or *required Q(on the condition that P)* (Fodor, 2000; Buller, 2005). For others, social contracts and precautions both belong to a more restricted class of deontic rules: those involving subjective utilities (e.g., Manktelow & Over, 1991). According to all of these theories, there should be no neural dissociations *within* the narrow domain of deontic rules involving utilities, that is, no dissociations between reasoning about social contracts versus precautionary rules. Yet there were. Because our results contradict the most domain-specific of the domain-general alternative hypotheses (deontic+utilities), they also contradict all domain-general hypotheses that include deontic rules involving utilities as a subset.

Inferences about the content of other people's mental states—TOM inferences—are necessary for interpreting rules involving social exchange but not for interpreting precautionary rules. That the computational requirements of each task differ in this way is supported by the neuroimaging results. Neural correlates of theory of mind (anterior and posterior temporal cortex) were differentially activated when subjects were interpreting social exchange scenarios, but not when they were interpreting precautionary ones. One TOM area (right parahippocampal gyrus at the amygdala) was activated when subjects were interpreting social rules describing people's preferences, habits or traits, when compared to activations for precautionary rules. In contrast to the interpretive phase, neural correlates of theory of mind were not activated for social contracts during the post-interpretive phase, during which subjects were deciding which individuals could have violated social contract or precautionary rules. Computing the desires of agents is logically necessary for interpreting a rule as involving social exchange. Once that

mapping has been made, cheater detection only requires that the mapping be remembered; it does not require further inferences from TOM.

Although detecting cheaters did not differentially activate TOM areas, detecting violations of precautionary rules produced a small activation in posterior temporal cortex (a neural correlate of TOM) along with large activations in a number of non-TOM brain areas. We do not have a specific interpretation of these particular precautionary activations, but the overall patterns during violation detection support the hypothesis that detecting cheaters, detecting people in danger, and detecting when people's preferences, habits or traits are inconsistent with a descriptive rule engage somewhat different neurocomputational machinery.

Deontic theories cannot be rescued by positing that processing social exchange and precautions differ *only* in that TOM inferences are activated while interpreting social exchange scenarios. Interpreting precautionary rules produced greater activation in many "non-TOM" areas of the brain, compared to interpreting social exchange rules; the same was true for the violation detection process (see Table 4). Moreover, at least one non-TOM area was more strongly activated by interpreting social exchange rules compared to precautionary ones (Table 3). If the same neurocomputational machinery processed all deontic rules, with the only difference being that TOM inferences were differentially engaged by social exchange, then we would not see activations in areas unrelated to theory of mind. Yet they occurred.

What about our conjecture about TOM(narrow) versus TOM(broad)? In their book, *Relevance: Communication and Cognition*, Sperber and Wilson (1995) provide an elegant analysis of communication as inference: Interpreting language requires inferences about the content of the speaker's mental states—inferences about what meaning the speaker intends to communicate. According to Sperber and colleagues, one subunit of the theory of mind system is a comprehension module, which evolved for inferring the communicative intent of speakers and treats linguistic utterances as metarepresentations (Sperber et al., 1995). In applying relevance theory to the Wason selection task, they posit that the comprehension module is equipped with procedures that spontaneously make logical inferences as well as ones that apply specific relevance principles (Sperber et al., 1995). Together, these procedures interpret conditional rules *without* engaging more domain-specific systems, such as social contract algorithms. According to their view, the comprehension module assigns the same logical form to deontic conditional rules, social contracts and precautions alike: *forbid(P and not-Q)*.

In response to this claim, Fiddick et al. (2000) present a number of behavioral results from Wason tasks involving social exchange that cannot be explained without invoking social contract algorithms and their domain-specific inferential rules. The neuroimaging results we report here support Fiddick et al.'s claim in two ways. First, during interpretation, neural correlates of theory of mind and at least one non-TOM area were activated by social exchange but *not* by precautionary rules. This result is difficult to understand if the same logical and relevance procedures are operating on and interpreting both types of rules. Second, a number of non-TOM areas were activated during interpretation of precautionary rules but *not* for social exchanges. Again, this finding suggests that the interpretive process is not identical for rules drawn from these two domains. These content-triggered dissociations are expected, however, if the comprehension module accesses a variety of domain-specific inference systems when interpreting the communicative intent of speakers: social contract algorithms, a domain-specific hazard-precaution system, as well as systems specialized for other forms of strategic interaction (e.g., aggressive threat (Tooby & Cosmides, 1989); coalitional cooperation (Tooby, Cosmides, & Price, 2006), anger as a negotiative system (Sell, 2005)).

Thus, as a friendly amendment to relevance theory, we suggest that a comprehension module would be better able to infer the content of speakers' mental states if it had access to all of these systems—to TOM(broad). Belief-desire inferences—TOM(narrow)—certainly feed into inferential systems that regulate strategic social interaction, like the social contract algorithms. But these inferential systems should also feed into TOM(narrow). The functional logic of social contract algorithms—and of other domain-specialized systems regulating strategic social interaction—can be used to infer the content of desires, goals, intentions, and beliefs (see above). Like the eye-direction detector (Baron-Cohen, 1995), we should expect social contract algorithms and other social inference systems to provide input for TOM(narrow).

Taken together, the operation of these interacting social inference systems would constitute the mind's "theory of human nature": TOHN. Belief-desire reasoning, TOM(narrow), would be a subunit of TOHN—one among many (Tooby, Cosmides & Price, 2006). A comprehension module equipped with TOHN would be a powerful inferential device, allowing people to negotiate the complex world of social interaction with a fuller understanding of other people's intentions.

Footnotes

1. We tried to avoid creating rules that can be interpreted as both a social contract *and* a precaution. E.g., drinkers would view "If you drink beer, you must be over 21 years old" as a social contract (it involves access to a benefit), whereas those making the rule view it as precautionary (drinking can lead to hazardous behavior); "If you play outside, you must wear your coat" is precautionary for the mother making the rule, but a social contract for the child who wants to play outside. To avoid such hybrid rules, we tried to make precautions in which the action in the antecedent was hazardous but *not* something people enjoy doing, and social contracts in which the consequent was not obviously precautionary.
2. Excluding the single social contract outlier gives Mann-Whitney $U = 7$, $p = .007$, consistent with the striking separation between problem sets one sees.

Acknowledgments

We thank David Turk for indispensable technical assistance and Tammy Laroche and Amy Rosenblum for recruiting subjects and running the scanner. We thank Howard Waldow and Mike Gazzaniga for making this project possible. This research was supported by a grant from the UCSB/Dartmouth Brain Imaging Program and an NIH Director's Pioneer Award to Leda Cosmides.

References

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

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience, 16*, 1773-1784.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442-481.

Atran, S. (1990). *The cognitive foundations of natural history*. New York: Cambridge University Press.

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

Baillargeon, R. (1987). Object permanence in 3.5- and 4.5-month-old infants. *Developmental Psychology, 23*, 655-664.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge: MIT Press.

Baron-Cohen, S. (2005). Autism and the origins of social neuroscience. In A. Easton & N. J. Emery (Eds.), *The cognitive neuroscience of social behaviour* (pp. 239-255). New York: Psychology Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a theory of mind? *Cognition, 21*, 37-46.

Baron-Cohen, S., Ring, H., A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., & Williams, S. C. R. (1999). Social intelligence in the normal and autistic brain: an fMRI study. *European Journal of Neuroscience, 11*, 1891-1898.

Barrett, H. C. (2005). Adaptations to predators and prey. In D. M. Buss (Ed.), *Handbook of Evolutionary Psychology* (pp. 220-223). Hoboken, NJ: John Wiley & Sons, Inc.

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

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

Boyer, P., & Lienard P. (in press). Why ritualized behavior? Precaution systems and action parsing in developmental, pathological, and cultural rituals. *Behavioral and Brain Sciences*.

Brace, G., Cosmides, L., & Tooby, J. (1998). Individuation, counting, and statistical inference: The role of frequency and whole object representations in judgment under uncertainty. *Journal of Experimental Psychology: General, 127*, 1-19.

Buller, D. J. (2005). *Adapting minds: Evolutionary psychology and the persistent quest for human nature*. Cambridge: MIT Press.

Calhoun, V. D., Stevens, M. C., Pearlson, G. D., & Kiehl, K. A. (2004). fMRI analysis with the general linear model: Removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *NeuroImage, 22*, 252-257.

Canessa, N., Gorini, A., Cappa, S. F., Piatelli-Palmarini, M., Danna, M., Fazio, F., & Perani, D. (2005). The effect of social content on deductive reasoning: An fMRI study. *Human Brain Mapping, 26*, 30-43.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12*, 1-47.

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

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

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257-271.

Cosmides, L. (1985). *Deduction or Darwinian algorithms? An explanation of the "elusive" content effect on the Wason selection task*. Doctoral dissertation, Department of Psychology, Harvard University: University Microfilms, #86-02206.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187-276.

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

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163-228). New York, NY: Oxford University Press.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*, 1-73.

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

Cosmides, L., & Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentation: A multidisciplinary perspective* (pp. 153-115). Vancouver Studies in Cognitive Science. New York: Oxford University Press.

Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Ed.), *Handbook of Evolutionary Psychology* (pp. 584-627). Hoboken, NJ: John Wiley & Sons, Inc.

Cosmides, L., Barrett, H. C., & Tooby, J. (forthcoming). Social contracts elicit the detection of intentional cheaters, not innocent mistakes.

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

Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping, 5*, 329-340.

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

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

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

Fiddick, L., Spampinato, M. V., & Grafman, J. (2005). Social contracts and precautions activate different neurological systems: An fMRI investigation of deontic reasoning. *NeuroImage, 28*, 778-786.

Fine, C., Lumsden, J., & Blair, R. J. R. (2001). Dissociation between "theory of mind" and executive functions in a patient with early left amygdala damage. *Brain, 124*, 287-298.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition, 57*, 109-128.

Fodor, J. (2000). Why we are so good at catching cheaters. *Cognition, 75,* 29-32.

Friston, K. J., Holmes, A. P., Worsely, K. J., Poline, J. B., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping, 2*, 189-210.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society London Series B, 358*, 459-473.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Science, 7*, 77-83.

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. (2000). Reading the mind in cartoons and stories: An fMRI study on "theory of mind" in verbal and nonverbal tasks. *Neuropsychologia, 38*, 11-21.

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

German, T. P., Niehaus, J. L., Roarty, M. P., Geisbrecht, B., & Miller, M. B. (2004). Neural correlates of detecting pretense: Automatic engagement of the intentional stance under cover conditions. *Journal of Cognitive Neuroscience, 16*, 1805-1817.

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

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gigerenzer, G. , Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gintis, H. (2000). *Game theory evolving*. Princeton, NJ: Princeton University Press.

Goel, V., & Dolan, R. J. (2001). Functional neuroanatomy of three-term relational reasoning. *Neuropsychologia, 39*, 901-909.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87*, B11-B22.

Goel, V., & Dolan, R. J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition, 93*, B109-B121.

Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImage, 12*, 504-514.

Goel, V., Gold, B., Kapur, S., & Houle, S. (1998). Neuroanatomical correlates of human reasoning. *Journal of Cognitive Neuroscience, 10*, 293-302.

Goel, V., Shuren, J., Sheesley, L., & Grafman, J. (2004). Asymmetrical involvement of the frontal lobes in social reasoning. *Brain, 127*, 783-790.

Henson, R. N. A., Rugg, M. D., & Friston, K. J. (2001). The choice of basis functions in event-related fMRI. *NeuroImage, 13*, S149.

Humphreys, N. K. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing Points in Ethology*. Cambridge: Cambridge University Press.

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

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94*, 412-426.

Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119-148). New York: Cambridge University Press.

Luce, L. R. (2003). Rationality in choice under certainty and uncertainty. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspective on judgment and decision research* (pp. 64-83). New York: Cambridge University Press.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697-720.

Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

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

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

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

Manktelow, K., & Over, D. (1988, July). Sentences, stories, scenarios, and the selection task. *First International Conference on Thinking*. Plymouth, UK.

Manktelow, K., & Over, D. (1990). Deontic thought and the selection task. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds.), *Lines of thinking* (Vol. 1, pp. 153-164). London: Wiley.

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

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

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*, 608-631 (1994).

Ollinger, J. M., Shulman, G. L., & Corbetta, M. (2001). Separating processes within a trial in event-related functional MRI. *NeuroImage, 15*, 547-558.

Panchanathan, K., & Boyd, R. (2003). A tale of two defectors: The importance of standing for the evolution of indirect reciprocity. *Journal of Theoretical Biology 224*, 115-126.

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

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

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55*, 87-124.

Saxe, R., & Kanwisher, N. (2003). People thinking about people: The role of the tempo-parietal junction in "theory of mind". *NeuroImage, 19*, 1835-1842.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right

temporo-parietal junction. *Neuropsychologia, 43*, 1391-1399.

Sell, A. (2005). *Regulating welfare tradeoff ratios: Three tests of an evolutionary-computational model of human anger*. Doctoral dissertation, Department of Psychology, University of California, Santa Barbara.

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M., & Rosen, B. R. (1997). Late onset of anterior prefrontal activity during true and false recognition: An event-related fMRI study. *NeuroImage, 6*, 259-269.

Spelke, E. (1990). Principles of object perception. *Cognitive Science, 14*, 29-56.

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

Sperber, D., & Girotto, V. (2002). Use or misuse of the selection task?: Rejoinder to Fiddick, Cosmides, and Tooby. *Cognition, 85*, 277-290.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Malden, MA: Blackwell Publishing.

Staddon, J. E. R. (1988). Learning as inference. In R. C. Bolles & M. D. Beecher (Eds.), *Evolution and learning* (pp. 59-77). Hillsdale, NJ: Erlbaum.

Stone, V. E., Baron-Cohen, S., Calder, A., Keane, J., & Young, A. (2003). Acquired theory of mind impairments in individuals with bilateral amygdala lesions. *Neuropsychologia, 41*, 209-220.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10*, 640-656.

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

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

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

Tooby, J., Cosmides, L., & Price, M. E. (2006). Cognitive adaptations for n-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics, 27*, 103-129.

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

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

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

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

Appendix

Text of Example Social Contract Problem: See Figure 3.

Text of Example Precaution Problem

Part 1: Tuberculosis (TB) is an airborne disease. You can get it from breathing in air that a TB patient has coughed or sneezed into. Nurses, who work with patients with all kinds of diseases, are advised: "If you work with patients with TB, then wear a surgical mask." You are wondering whether any of these nurses ever break this safety rule.

Part 2: You will see cards representing some nurses. Each card represents one nurse. One side of the card tells whether or not that nurse worked with TB patients on a particular day, and the other side tells whether or not that nurse wore a surgical mask that day. You are concerned that some of these nurses may be in danger.

Part 3: As you see each card, tell us if you would definitely need to turn over that card to find out if that nurse has violated the rule: "If you work with patients with TB, then wear a surgical mask." Don't turn over any more cards than are absolutely necessary.

Card (not-Q): Could this person have violated the rule? Card: "Lindsey did not wear a surgical mask". Rule: "If you work with patients with TB, then wear a surgical mask."

Text of Example Descriptive Problem

Part 1: Sometimes it seems that people who go into a profession are similar in certain ways. Your friend Bill says he has been watching accountants, forest rangers, lawyers, and biologists, and has noticed the following rule holds: "If a person becomes a biologist, then that person enjoys camping." You want to see whether people's preferences ever violate this rule.

Part 2: You will see cards representing some people. Each card represents one person. One side of the card tells whether or not that person is a biologist, and the other side tells whether or not that person enjoys camping. You are concerned that Bill's rule may be wrong.

Part 3: As you see each card, tell us if you would definitely need to turn over that card to find out if that case violates the rule: "If a person becomes a biologist, then that person enjoys camping." Don't turn over any more cards than are absolutely necessary.

Card (Q card): Could this person have violated the rule? Card: Paul enjoys camping. Rule: "If a person becomes a biologist, then that person enjoys camping."

Table 1

Brain areas activated during stories (collpased across content) compared to rest (fixation cross).

Brain Area	Hemi	BA	Voxels	x	y	z	t value
Stories > Rest							
Superior Frontal Gyrus	L	10	34	-6	68	27	5.67
	L	8	57	-6	20	49	5.91
	L	8	12	-36	17	54	4.92
Middle Frontal Gyrus	L	46	50	-56	33	15	7.38
	L	46	13	-39	27	18	4.28
	L	9	43	-42	8	36	4.29
Superior Temporal Gyrus	L	39	24	-53	-57	28	4.91
Middle Temporal Gyrus	L	21	381	-53	-27	-6	7.82
Angular Gyrus	L	39	42	-30	-59	39	4.83
Superior Parietal Lobule	L	7	18	-33	-67	56	4.76
Fusiform Gyrus	L	19	1187	-27	-77	-19	7.68
Rest > Stories							
Superior Frontal Gyrus	L	11	20	-27	43	-15	4.92
	R	8	10	24	49	39	4.75
	R	8	23	21	20	46	6.78
	R	8	12	18	40	50	5.54
Medial Frontal Gyrus	L	10	1279	-9	52	0	13.22
Middle Frontal Gyrus	L	9	34	-30	36	26	5.05
	R	47	74	45	35	-4	5.85
Inferior Frontal Gyrus	R	44	23	50	13	19	5.88
Middle Temporal Gyrus	L	21	592	-42	-3	-7	7.29
	R	39	13	45	-58	8	4.19
	R	39	52	45	-63	28	5.51
Inferior Parietal Lobule	R	40	149	56	-31	29	9.78
Superior Occipital Gyrus	L	19	47	-45	-80	37	5.20
Cingulate Gyrus	L	31	1674	-9	-24	40	10.31
Parahippocampal Gyrus	L	34	27	-9	-10	-20	4.77
Insula	R	13	29	33	-22	18	4.41
Caudate Body	L	--	10	-15	18	13	4.60
Caudate Tail	L	--	29	-21	-34	18	6.74
	L	--	13	-30	-43	10	6.23
	R	--	12	24	-43	10	4.20

Note: Hemi = hemisphere, L = left, R = right. BA = Broadmann's area based on stereotaxic coordinates. x, y, z values are Talairach coordinates. Statistical threshold: $p < .005$, extent = 10 voxels.

Table 2

Brain areas activated during card responses (collpased across content) compared to rest (fixation cross).

Brain Area	Hemi	BA	Voxels	X	y	z	t value
Cards > Rest							
Superior Frontal Gyrus	L	8	208	-3	20	49	9.33
Medial Frontal Gyrus	R	9	16	9	31	34	4.99
Middle Frontal Gyrus	L	9	412	-48	16	32	12.28
	R	9	97	53	28	29	7.12
	L	10	45	-36	56	19	5.63
	R	10	21	33	48	20	6.94
Middle Temporal Gyrus	L	20	276	-50	-35	-6	6.30
	R	21	10	56	-30	-9	4.56
Fusiform Gyrus	L	37	16	-39	-42	-21	5.40
Inferior Parietal Lobule	L	40	354	-42	-53	47	8.63
	R	40	110	36	-50	41	9.08
Lingual Gyrus	L	17	573	-15	-94	-13	14.76
	R	18	276	24	-94	-5	10.26
Lentiform Nucleus Putamen	L	--	35	-21	3	8	6.37
Cerebellum	R	--	243	48	-65	-24	6.53
Rest > Cards							
Superior Frontal Gyrus	R	8	21	21	23	49	5.80
	R	8	21	18	40	50	6.51
Medial Frontal Gyrus	L	10	791	-3	46	-7	14.01
Middle Frontal Gyrus	L	8	41	-24	31	37	4.30
Precentral Gyrus	R	6	14	53	1	11	5.43
Paracentral Lobule	L	31	7417	0	-27	46	16.31
Superior Temporal Gyrus	R	38	13	36	13	-41	5.42
Middle Temporal Gyrus	L	39	207	-42	-74	29	11.01
Angular Gyrus	R	39	150	48	-69	28	7.67
Cuneus	L	19	56	-15	-83	35	5.58

Note: Hemi = hemisphere, L = left, R = right. BA = Broadmann's area based on stereotaxic coordinates. x, y, z values are Talairach coordinates. Statistical threshold: $p < .005$, extent = 10 voxels.

Table 3

Brain areas activated for social contracts.

Brain Area	Hemi	BA	Voxels	x	y	z	t value
Stories							
Social Contracts > Precautions							
Precentral Gyrus	R	6	11	65	3	5	4.14
Middle Temporal Gyrus	L	21	14	-68	-38	-6	4.91
Inferior Temporal Gyrus	R	20	16	62	-10	-22	5.44
Posterior Cingulate	R	23	11	3	-63	14	4.79
Social Contracts > Descriptives							
Superior Temporal Gyrus	L	22	20	-50	-15	1	5.18
	R	22	13	48	8	-5	7.36
Cards							
Social Contracts > Precautions							
No clusters survived threshold.							
Social Contracts > Descriptives							
No clusters survived threshold.							

Note: Hemi = hemisphere, L = left, R = right. BA = Broadmann's area based on stereotaxic coordinates. x, y, z values are Talairach coordinates. Statistical threshold: $p < .005$, extent = 10 voxels.

Table 4

Brain areas activated for precautions.

Brain Area	Hemi	BA	Voxels	x	y	z	t value
Stories							
Precautions > Social Contracts							
Superior Frontal Gyrus	L	9	12	-15	42	31	6.25
Medial Frontal Gyrus	L	6	12	-18	-9	50	4.68
Postcentral Gyrus	L	2	17	-30	-22	31	5.48
Cingulate Gyrus	R	31	67	24	-30	35	6.89
	R	31	63	21	19	32	5.59
Brainstem Pons	L	--	23	-6	-16	-27	5.38
Precautions > Descriptives							
Superior Frontal Gyrus	L	10	10	-24	55	-3	4.82
Medial Frontal Gyrus	R	10	16	15	56	6	6.58
Inferior Parietal Lobule	R	40	31	62	-30	37	5.78
Precuneus	R	7	21	15	-38	49	5.52
Cingulate Gyrus	R	31	11	9	-24	43	4.99
Cards							
Precautions > Social Contracts							
Superior Frontal Gyrus	L	6	21	-24	11	49	9.47
	R	10	14	27	47	0	4.36
Middle Frontal Gyrus	R	46	10	42	44	6	4.48
	R	46	11	42	18	18	5.41
	R	46	29	42	36	18	7.74
	R	6	13	36	14	52	3.93
Inferior Frontal Gyrus	L	47	12	-42	17	-6	4.80
Superior Temporal Gyrus	L	41	12	-36	-29	7	4.85
Middle Temporal Gyrus	L	21	14	-59	-26	-1	6.59
Fusiform Gyrus	L	37	21	-30	-50	-10	4.99
Cuneus	R	18	16	3	-77	26	5.58
Cingulate Gyrus	R	24	32	12	-4	33	5.15
Insula	R	13	37	42	-31	18	4.22
	R	13	18	33	21	7	4.62
Cerebellum	R	--	61	12	-65	-12	6.57
Precautions > Descriptives							
Precentral Gyrus	L	6	16	-62	-18	45	4.88
	R	4	19	21	-23	73	5.90
Postcentral Gyrus	R	3	12	42	-26	65	6.70

Note: Hemi = hemisphere, L = left, R = right. BA = Broadmann's area based on stereotaxic coordinates. x, y, z values are Talairach coordinates. Statistical threshold: $p < .005$, extent = 10 voxels.

Table 5

Brain areas activated for descriptives.

Brain Area	Hemi	BA	Voxels	x	y	z	t value
Stories							
Descriptives > Social Contracts							
No clusters survived threshold.							
Descriptives > Precautions							
Parahippocampal Gyrus Amygdala	R	--	13	24	-1	-15	5.21
Cards							
Descriptives > Social Contracts							
Superior Frontal Gyrus	R	8	29	3	32	51	5.02
Medial Frontal Gyrus	L	9	36	-3	39	28	5.48
Precentral Gyrus	L	9	19	-39	16	38	4.27
	R	9	22	45	22	35	4.91
Fusiform Gyrus	R	18	28	24	-86	-21	4.72
Cingulate Gyrus	L	24	43	0	-7	25	5.22
	L	31	10	-3	-45	27	4.29
Cerebellum	L	--	13	-12	-37	-49	5.22
Descriptives > Precautions							
Middle Frontal Gyrus	L	46	11	-50	27	24	4.55
	R	9	30	45	25	35	5.43
Inferior Parietal Lobule	L	7	18	-36	-59	47	4.58

Note: Hemi = hemisphere, L = left, R = right. BA = Broadmann's area based on stereotaxic coordinates. x, y, z values are Talairach coordinates. Statistical threshold: $p < .005$, extent = 10 voxels.

Figure Captions

Figure 1. Which domain is the system that causes reasoning about social contracts designed for? The hypothesis tested herein is that this system is designed to operate on social contracts. Alternative hypotheses hold that it is designed to operate on all deontic conditional rules involving utilities; all deontic conditional rules; all familiar conditionals; or all conditionals. Social contracts belong to each of these more general categories. If reasoning about social contracts dissociates from reasoning about precautions—that is, if there is a dissociation within the class of deontic rules involving utilities—then all the more domain general alternatives fail.

Figure 2. The Wason selection task: Syntax of social contract (A), precautionary (B), and descriptive (C) problems. They all have the same logical structure: If P then Q. They differ only in content (i.e., what P and Q stand for): social contracts specify benefits that are conditional on meeting the provisioner's requirement whereas precautionary rules specify hazardous activities that can be made safer by taking an appropriate precaution. Check marks indicate correct card choices. On these problems, looking for cheaters and looking for people in danger results in choosing the logically correct cards.

Figure 3. Illustration of screen displays seen by subjects when reasoning about a social contract problem (not to scale). Story is shown in panels A-C, cards in panels E and G. Two versions of the P card are shown. For each story, subjects saw a total of eight cards, two versions of each logical category.

Figure 4. Activations in story contrasts overlayed on a 3D rendering of a mean anatomical image ($p < .005$, uncorrected, extent = 10 voxels). Top panel shows significant clusters for the social contract > precaution story contrast in red; significant clusters for the reverse contrast (precaution > social contract) are shown in blue. Middle panel shows significant clusters for the social contract > descriptive story contrast, and bottom panel for the descriptive > precaution story contrast.

Figure 5. Anterior (top panels) and posterior (bottom panels) temporal lobe clusters significantly more active for social contract stories compared to precaution stories overlayed on a mean anatomical image ($p < .005$, uncorrected, extent = 10 voxels). Graphs show the time course of the BOLD signal. The flat line corresponds to the stasis assumed by our model (i.e., the "box car" portion of the model). Plots were obtained by averaging parameters estimated by the model across all voxels in the cluster.

Figure 6. The pattern of differential activation replicates across individual problems, supporting the hypothesis that these activation differences are driven by the underlying, content-specific representation of social exchange versus precautionary problems. The average signal intensity for each individual social contract and precautionary problem is shown for three brain areas in which there was greater activation for social contract than precautionary problems: (A) anterior temporal cortex (BA 20; x = 62, y = -10, z = -22); (B) middle temporal cortex (BA 21; x = -68, y = -38, z = -6); (C) posterior cingulate (BA 23; x = 3, y = -63, z = 14). There is very little overlap between the problem sets.

Figure 1

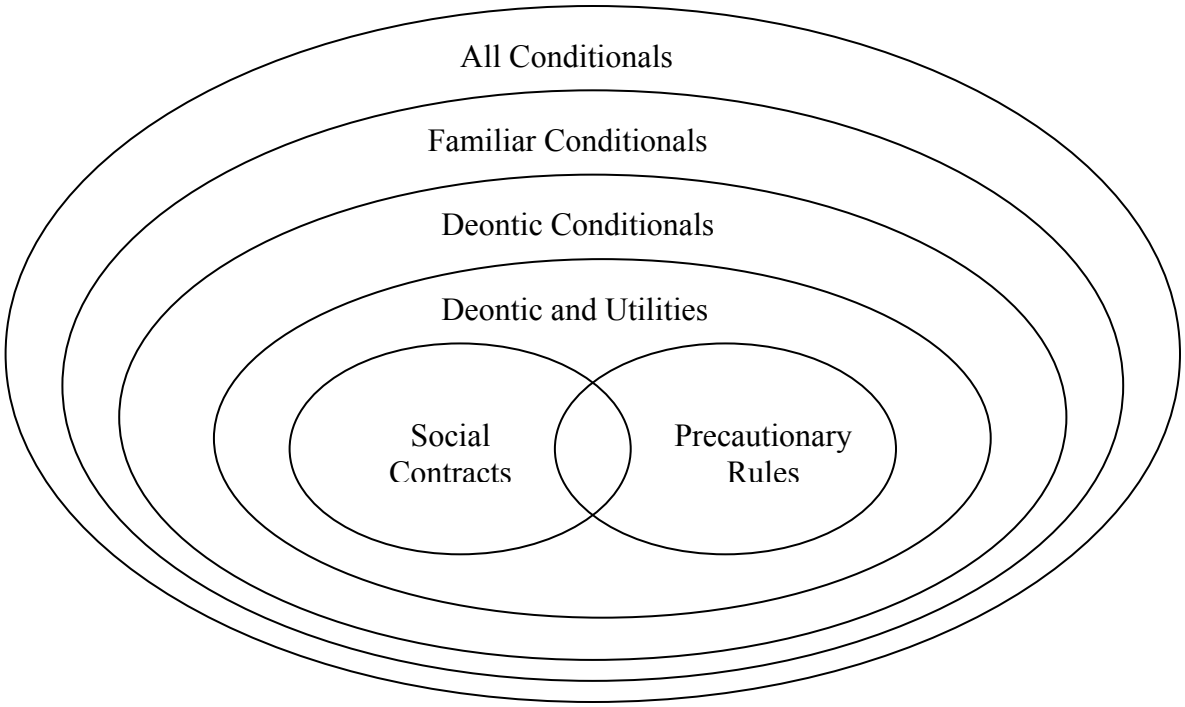
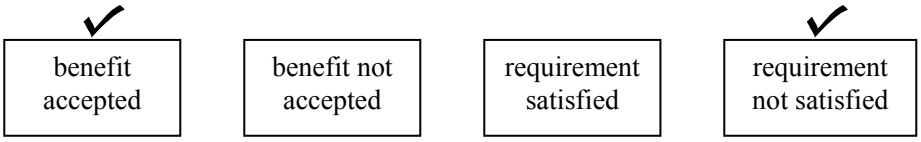

All Conditionals
Familiar Conditionals
Deontic Conditionals
Deontic and Utilities
Social Contracts
Precautionary Rules

Figure 2

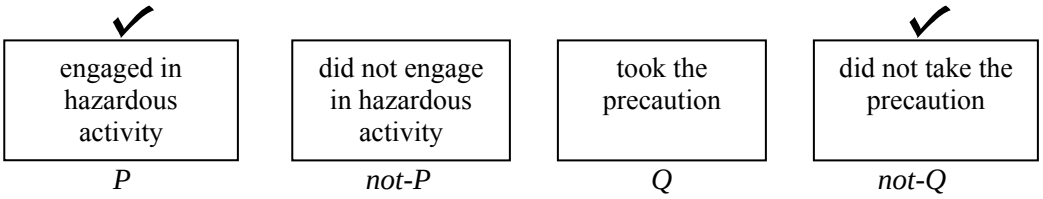
A. Syntax of a Social Contract Problem

The following rule holds: **If you take the benefit, then you must satisfy the requirement.**
(*If* *P* *then* *Q*)
You want to see whether anyone ever violates this rule. The cards below have information about four people. Each card represents one person. One side of the card tells whether or not that person accepted the benefit, and the other side tells whether or not that person satisfied the requirement. You are concerned that someone may have violated the rule. Indicate which card(s) you would definitely need to turn over to see if any of these people have violated the rule.



B. Syntax of a Precautionary Problem

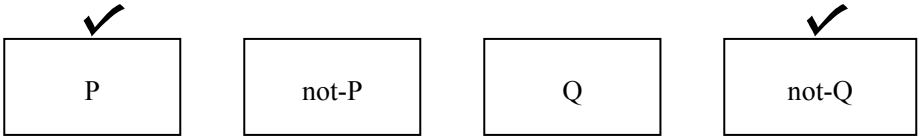
The following rule holds: **If you engage in the hazardous activity, then you must take the precaution.**
(If P then Q)
You want to see whether anyone ever violates this rule. The cards below have information about four people. Each card represents one person. One side of the card tells whether or not that person is engaging in the hazardous activity, and the other side tells whether or not that person has taken the precaution. You are concerned that someone may have violated the rule. Indicate which card(s) you would definitely need to turn over to see if any of these people have violated the rule.



C. Syntax of a Descriptive Problem (indicative conditional rule, social content)

You've been told the following rule holds:
If a person is in category P, then that person has preference [or habit or trait] Q.

You want to see whether people's preferences ever violate this rule. The cards below have information about four people. Each card represents one person. One side of the card tells whether or not that person is in category P, and the other side tells whether or not that person has preference Q. You are concerned that the rule may be wrong. Indicate which card(s) you would definitely need to turn over to see if any of these people's preferences violate the rule.



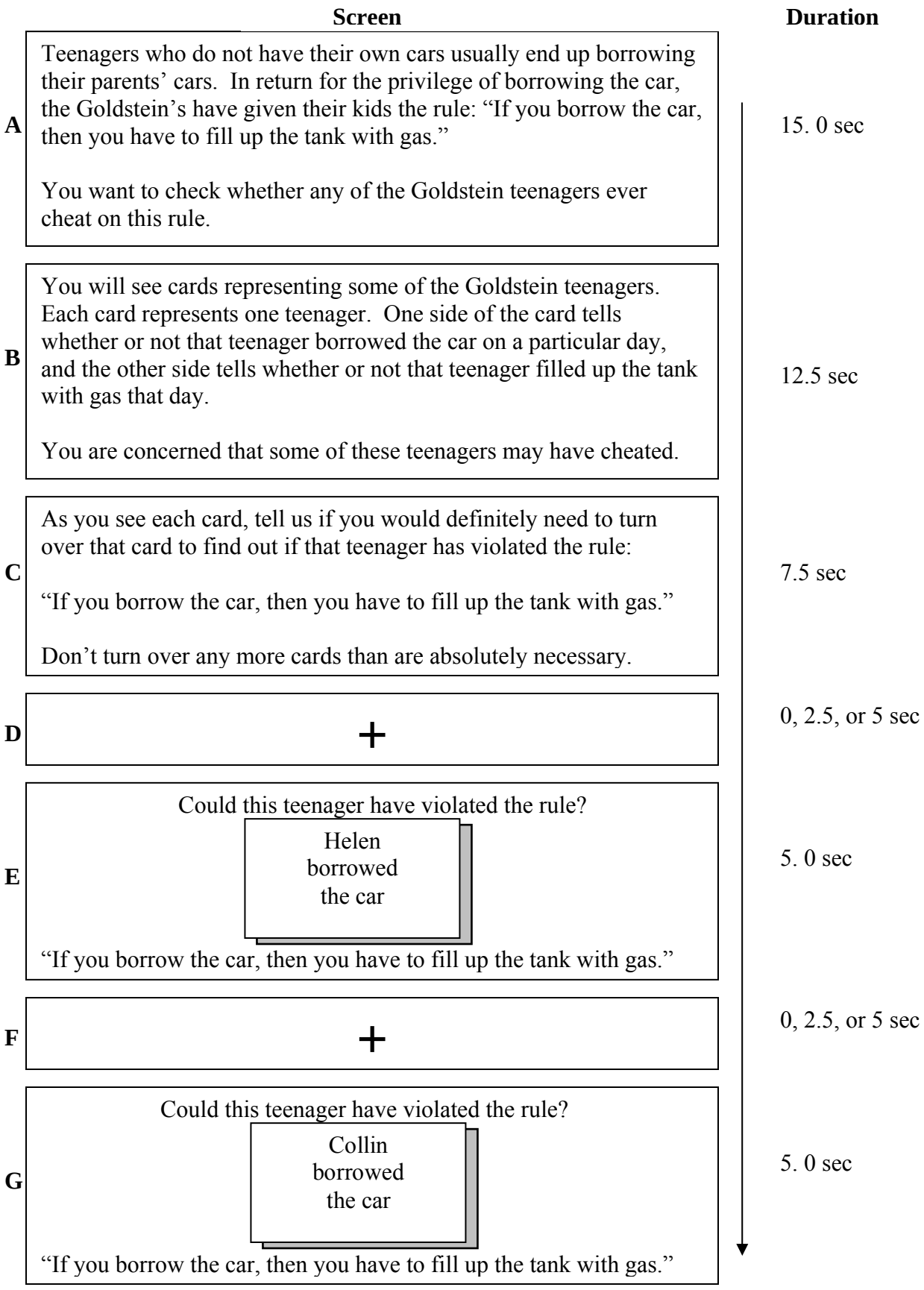

Screen
Duration
A
Teenagers who do not have their own cars usually end up borrowing their parents' cars. In return for the privilege of borrowing the car, the Goldstein's have given their kids the rule: "If you borrow the car, then you have to fill up the tank with gas."
You want to check whether any of the Goldstein teenagers ever cheat on this rule.
15. 0 sec
B
You will see cards representing some of the Goldstein teenagers. Each card represents one teenager. One side of the card tells whether or not that teenager borrowed the car on a particular day, and the other side tells whether or not that teenager filled up the tank with gas that day.
You are concerned that some of these teenagers may have cheated.
12.5 sec
C
As you see each card, tell us if you would definitely need to turn over that card to find out if that teenager has violated the rule:
"If you borrow the car, then you have to fill up the tank with gas."
Don't turn over any more cards than are absolutely necessary.
7.5 sec
D
+
0, 2.5, or 5 sec
E
Could this teenager have violated the rule?
Helen borrowed the car
"If you borrow the car, then you have to fill up the tank with gas."
5. 0 sec
F
+
0, 2.5, or 5 sec
G
Could this teenager have violated the rule?
Collin borrowed the car
"If you borrow the car, then you have to fill up the tank with gas."
5. 0 sec

Figure 4

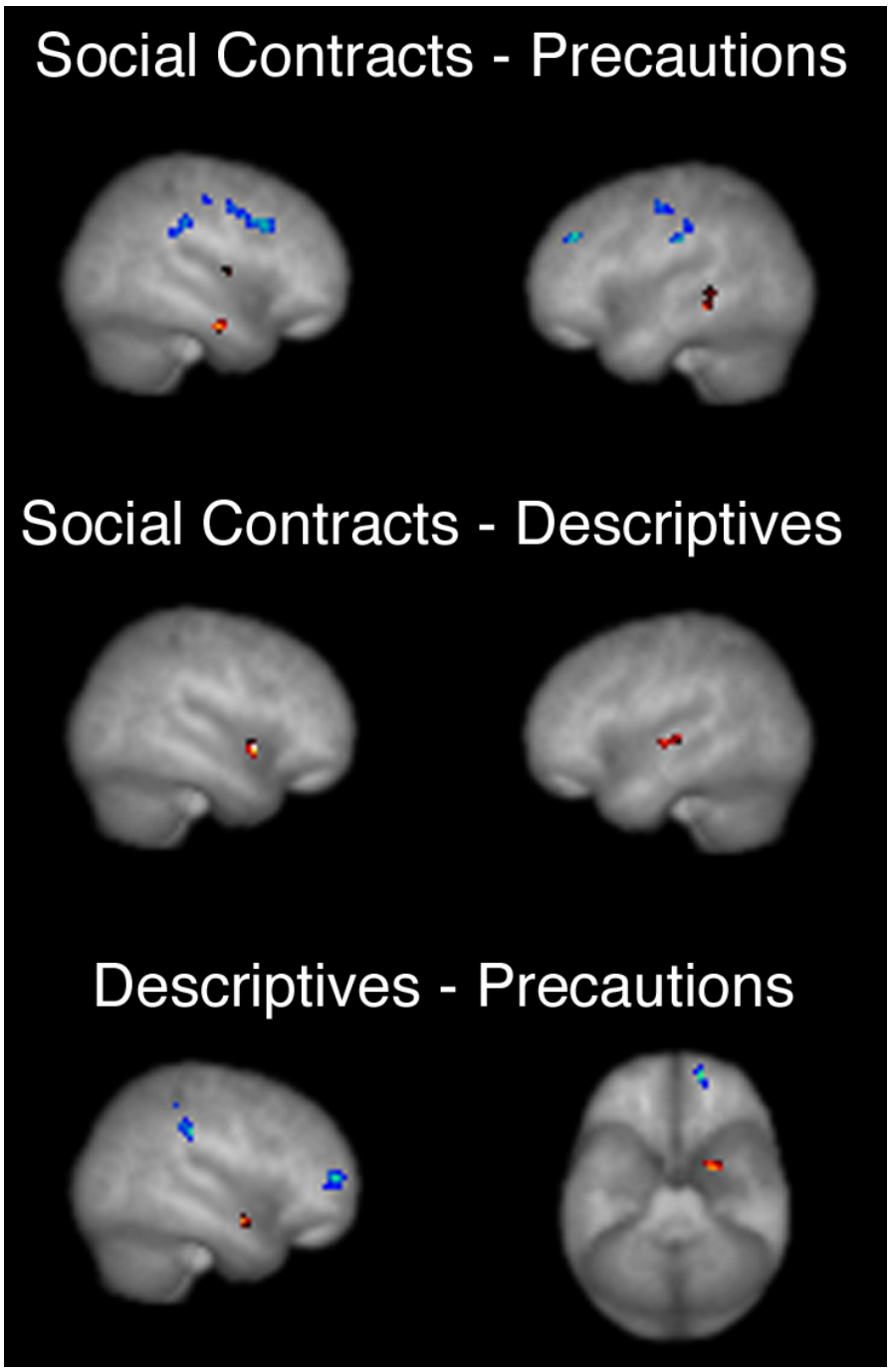

Social Contracts - Precautions
Social Contracts - Descriptives
Descriptives - Precautions

Figure 5

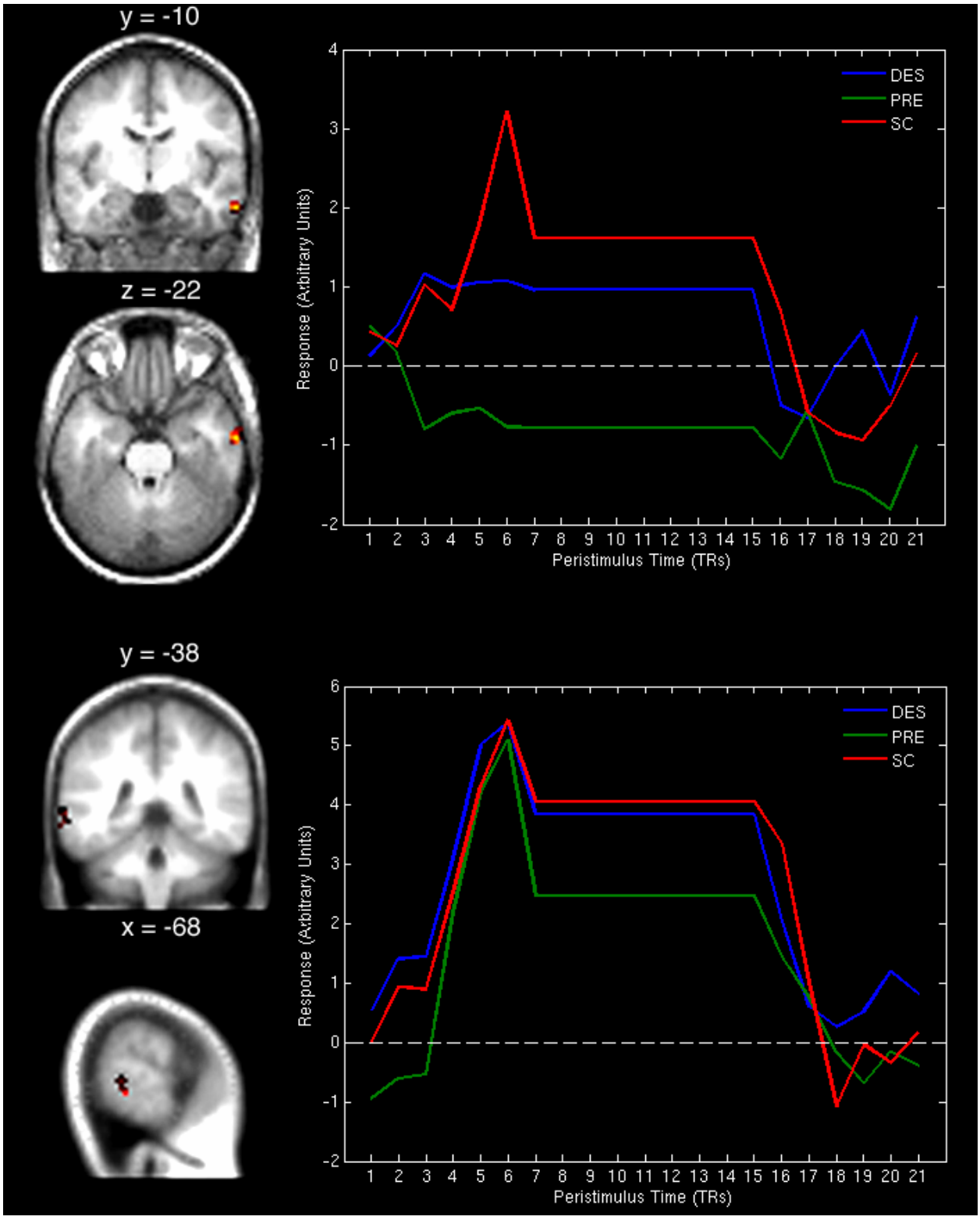
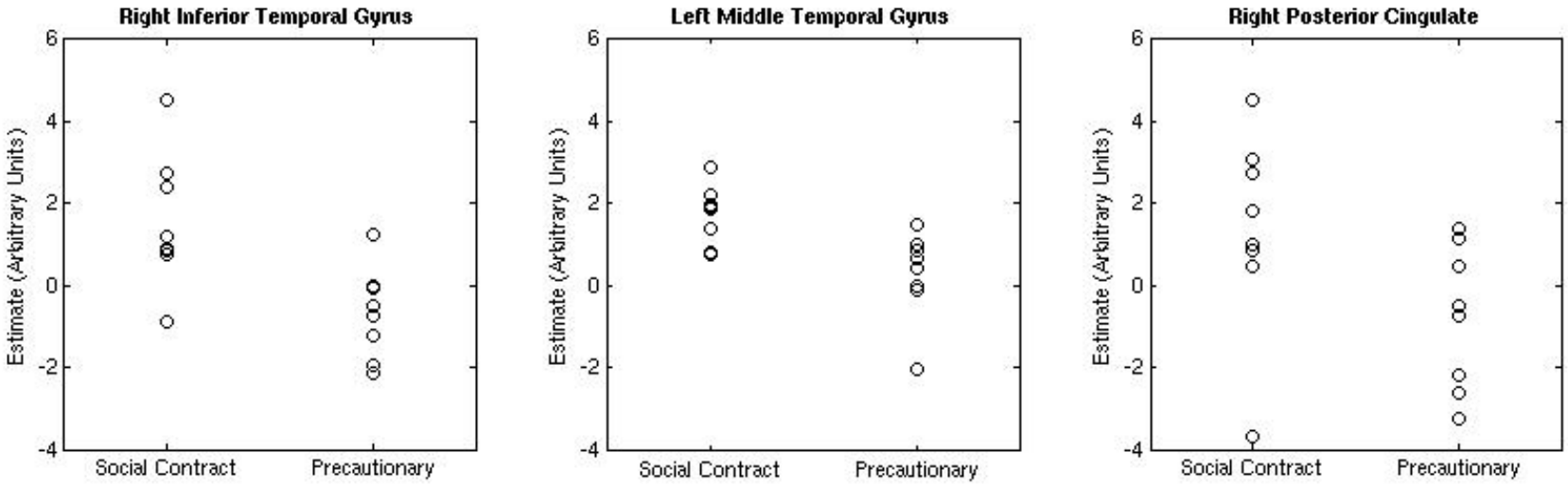

y = -10
z = -22
y = -38
x = -68
DES
PRE
SC
Response (Arbitrary Units)
Peristimulus Time (TRs)
DES
PRE
SC
Response (Arbitrary Units)
Peristimulus Time (TRs)

Figure 6


Right Inferior Temporal Gyrus
Left Middle Temporal Gyrus
Right Posterior Cingulate
Estimate (Arbitrary Units)
6
4
2
0
-2
-4
Social Contract
Precautionary