ELSEVIER

Cognition 77 (2000) 1–79

COGNITION

www.elsevier.com/locate/cognit

No interpretation without representation: the role of domain-specific representations and inferences in the Wason selection task

Laurence Fiddick[a], Leda Cosmides[b,*], John Tooby[c]

[a]*Max-Planck-Institute for Human Development, Center for Adaptive Behavior and Cognition, Lentzeallee 94, D-14195 Berlin, Germany*
[b]*Department of Psychology, University of California, Santa Barbara, CA 93106, USA*
[c]*Center for Evolutionary Psychology, Department of Anthropology, University of California, Santa Barbara, CA, USA*

Received 20 December 1997; accepted 15 April 2000

Abstract

The Wason selection task is a tool used to study reasoning about conditional rules. Performance on this task changes systematically when one varies its content, and these content effects have been used to argue that the human cognitive architecture contains a number of domain-specific representation and inference systems, such as social contract algorithms and hazard management systems. Recently, however, Sperber, Cara & Girotto (Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, *57*, 31–95) have proposed that relevance theory can explain performance on the selection task – including *all* content effects – without invoking inference systems that are content-specialized. Herein, we show that relevance theory alone cannot explain a variety of content effects – effects that were predicted in advance and are parsimoniously explained by theories that invoke domain-specific algorithms for representing and making inferences about (i) social contracts and (ii) reducing risk in hazardous situations. Moreover, although Sperber et al. (1995) were able to use relevance theory to produce some new content effects in other domains, they conducted no experiments involving social exchanges or precautions, and so were unable to determine which – content-specialized algorithms or relevance effects – dominate reasoning when the two conflict. When experiments, reported herein, are constructed so that the different theories predict divergent outcomes, the results support the predictions of social contract theory and hazard management theory, indicating that these inference systems override content-general relevance factors. The fact that social contract and hazard management algorithms provide

* Corresponding author. Fax: +1-805-893-4303.
E-mail addresses: cosmides@psych.uscb.edu (L. Cosmides), fiddick@mpib.berlin.mpg.de (L. Fiddick), tooby@sscf.ucsb.edu (J. Tooby).

PII: S0010-0277(00)00085-8

better explanations for performance in their respective domains does not mean that the content-general logical procedures posited by relevance theory do not exist, or that relevance effects never occur. It does mean, however, that one needs a principled way of explaining which effects will dominate when a set of inputs activate more than one reasoning system. We propose the *principle of pre-emptive specificity* – that the human cognitive architecture should be designed so that more specialized inference systems pre-empt more general ones whenever the stimuli centrally fit the input conditions of the more specialized system. This principle follows from evolutionary and computational considerations that are common to both relevance theory and the ecological rationality approach.

Keywords: Reasoning; Relevance; Social exchange; Social contracts; Cooperation; Logic; Evolution; Evolutionary psychology

1. Introduction

What is the nature of the computational machinery that causes human reasoning? In the past decade, research in cognitive development, cognitive neuroscience, and evolutionary psychology has been converging on the view that the human cognitive architecture was designed by natural selection to reliably develop a number of expert systems (for reviews, see Barkow, Cosmides & Tooby, 1992; Hirschfeld & Gelman, 1994; Pinker, 1997; Sperber, 1994). Each is equipped with (i) a proprietary format that represents distinctions that were adaptively important in the domain for which it evolved, and (ii) functionally-specialized inferential procedures that were designed to operate on these proprietary representations, generating inferences that, while not true across domains, were adaptively useful when operating within the system's proper domain of application. Examples include the 'mindreading' system (e.g. Baron-Cohen, 1995; Leslie, 1987), an object mechanics system (e.g. Baillargeon, 1986; Leslie & Keeble, 1987; Spelke, 1990), a system for inducing grammar (Pinker, 1994), and a system for understanding the biological world (Caramazza & Shelton, 1998; Gutheil, Vera & Keil, 1998; Hatano & Inagaki, 1994; Keil, 1994). We have proposed several expert systems of this kind, two of which we will discuss here: one designed for reasoning about social exchange (Cosmides, 1985, 1989; Cosmides & Tooby, 1989, 1992) and another designed for reasoning about risk reduction in hazardous situations (Cosmides & Tooby, 1997; Fiddick, 1998; Fiddick, Cosmides & Tooby, 1995).

A central method used to test domain-specific reasoning theories has been to see whether the inferences people make vary as a function of the content they are asked to reason about (see Hirschfeld & Gelman, 1994). The discovery of such content effects is required to support these theories, but one must always ask whether these same effects can be better explained by a theory that does not invoke content-specialized inferential machinery.

A great deal of data used to support domain-specific theories of reasoning (includ-

ing our own) has been generated with the Wason selection task. This paper-and-pencil test is one of the most extensively used tools in the study of reasoning, and has properties that make it well-suited for exploring content effects. Recently, however, Sperber, Cara and Girotto (1995) claim to have explained performance on this task – including *all* content effects – without having to invoke content-specialized computational machinery. They do this by positing an evolved module that is itself domain-specific – it is specialized for the domain of discourse comprehension – although not content-specialized. They maintain that a further implication of their view is that the selection task should be abandoned as a means of investigating reasoning because it cannot illuminate the underlying computations. If they are correct in either claim, then this would have significance for how a large community of cognitive scientists should proceed. We will argue that, while Sperber et al. (1995) (henceforth SCG) have produced an insightful theory that is well worth further exploration, neither of these larger claims is supported by the evidence.

1.1. The Wason selection task

Developed by Peter Wason in the 1960s (Wason, 1966, 1968), the selection task first achieved notoriety by inciting a debate about whether the human cognitive architecture implements rules of logical inference (which is still unresolved; see Bonatti, 1994; Johnson-Laird & Byrne, 1991; Rips, 1994; Wason & Johnson-Laird, 1972). Its structure is simple. The subject is given a conditional rule of the form *If P then Q*, and shown (pictures of) four cards with information pertaining to the rule. Each card represents a separate instance that might satisfy or violate the rule; one side tells whether that instance has property *P* and the other side tells whether it has property *Q*. The subject can see only one side of each card, and the sides facing the subject display the values *P*, *not-P*, *Q*, and *not-Q* (indicating, respectively, an instance that has property *P*, one that lacks it (*not-P*), one that has property *Q*, and one that lacks it (*not-Q*)). Subjects are then asked which of the four cards they would need to turn over to see whether any of these instances violate the rule.[1]

A conditional rule is violated by any instance that has property *P* but lacks property *Q*: that is, by instances of *P-and-(not-Q)*, a true antecedent paired with a false consequent. The seemingly straightforward solution to the task is to investigate instances for which *P* is true and instances for which *Q* is false, as this will reveal any possible cases in which the feature *P* is conjoined with the feature *not-Q*. The remarkable thing about the task is that, despite its apparent simplicity, subjects routinely fail to perform according to the norms of formal logic.[2] Instead of selecting the *P* card, the *not-Q* card, and no others, most subjects select the *P* and *Q* cards, or the *P* card alone (reviewed in Evans, 1982). Yet the task is conceptually so simple,

[1] There are variations (e.g. subjects are sometimes asked which cards need to be turned over to determine whether the rule is true), but the violation wording given here is the least ambiguous: it does not require that the *rule's* truth be evaluated, and has a clear normative answer.

[2] Subjects fail to give the logically correct answer, whether one interprets 'If P then Q' as a material conditional or as a biconditional (i.e. as also implying 'If Q then P'). On the biconditional interpretation, the logically correct answer is to choose all four cards. Yet very few subjects typically give this response.

the solution could be implemented in a single line of code if conditional rules were spontaneously represented according to their logical form.

This failure to answer '*P & not-Q*', which was first noted with descriptive rules (e.g. 'If a person eats hot chili peppers, then he will drink a cold beer'), was initially thought to apply to all conditional rules. In the last decade, however, researchers have been successful in eliciting *P* and *not-Q* selections with tasks employing a deontic conditional – a conditional statement describing what a person is *obligated* or *entitled* to do in a given context (reviewed in Manktelow & Over, 1995). There is still considerable debate over whether these results are best accounted for by mechanisms whose scope approximates the entire domain of deontic rules (Cheng & Holyoak, 1985; Cheng, Holyoak, Nisbett & Oliver, 1986; Manktelow & Over, 1990, 1991) or by a series of more domain-specific competences (Cosmides, 1985, 1989; Cosmides & Tooby, 1992, 1997; Cummins, 1996, 1998; Fiddick, 1998; Fiddick, Cosmides & Tooby, 2000; Gigerenzer & Hug, 1992; Platt & Griggs, 1993; Stone, Cosmides, Tooby, Knight & Kroll, 2000; Sugiyama, Tooby & Cosmides, 2000), but we will not focus on these differences here. What is common to these accounts is that they posit specialized mechanisms for reasoning within restricted content domains.

Seemingly opposed to these domain- and content-specific accounts of content effects on the selection task is a recent proposal by SCG that "Relevance theory explains the selection task" (their article's title). They maintain that performance is heavily influenced by how subjects interpret the rule in the context of the task. They then argue that interpretations are triggered by an interplay of logical inferences and content-general principles of relevance. To underline the content-generality of the procedures they proposed, they demonstrated in several cases that logical performance can be elicited by selection tasks employing non-deontic, descriptive rules, given the right pragmatic context. Although they neither devised nor tested any deontic rules experimentally, they went on to assert that theirs is a "wholly general" and comprehensive account of existing content effects on the selection task (SCG, p. 88). They maintain that they have provided an alternative explanation, to be preferred on grounds of generality, of past results on deontic problems that have been widely interpreted as support for content-specific mechanisms.

As Sperber and his colleagues make clear (Sperber, 1994; Sperber & Wilson, 1986, 1995; SCG), their proposal is not intended to deny that evolved domain-specific inference mechanisms, such as social contract algorithms, exist, or that they play an important role in mental life. Indeed, Sperber has weighed in powerfully in this debate in favor of the necessity of evolved, content-specialized mechanisms in many inferential contexts (e.g. Sperber, 1994). The authors simply argue that, in explaining content effects *on the Wason selection task*, once relevance effects are subtracted, either (1) there is nothing else to explain, or (2) relevance effects inevitably confound experiments, rendering the Wason task useless as a means of detecting the operation of content-specialized inference mechanisms. They suggest that, given their results, the selection task will have to be abandoned as a method for demonstrating the existence or exploring the properties of domain-specific inference mechanisms.

We strongly endorse the proposition that multiple independent converging lines of evidence ought to be used to test for the existence and properties of any cognitive device; on social exchange, for example, see the experimental economics work by Hoffman, McCabe and Smith (1998) and Kurzban, Rutherford, Cosmides and Tooby (1997) and the cognitive neuroscience studies of Stone, Cosmides and Tooby (1996) and Stone et al. (2000). We also think that relevance theory, as SCG have applied it, potentially illuminates performance on selection tasks employing descriptive (i.e. non-deontic, indicative) rules. The area of disagreement involves relevance theory's claims to have provided a comprehensive explanation for all content effects through a single, overarching general theory.

There is an intimate relationship between how a situation is represented, and what inferences will be made about it. This is because inferential procedures are defined by how they operate over particular categories of representation. For this reason, every theory of interpretation presupposes a representational system in which the interpretations are formulated and inferences are made. SCG's theory of interpretation as applied to the Wason selection task employs logical representations and inference procedures, which are content-independent. But when scrutinized, these turn out to be far too impoverished to account for the corpus of data on deontic rules: the inferences subjects draw cannot be derived by the application of logical rules alone. As we will show, when a situation is represented as involving social exchange or hazard management, content-specific inferential machinery is activated, which regulates both the interpretive process and post-interpretive inferences.

More specifically, we will argue as follows:

1. A careful application of relevance theory to subject performance on deontic rules shows that it does not and cannot explain many of the most striking existing results. Therefore, it does not provide a more general, alternative explanation for deontic performance.
2. To explain the highly patterned functional specificity of these results, one is forced to invoke content-specialized inferential machinery, including social contract algorithms and hazard management algorithms.
3. For straightforward evolutionary functional reasons, the human cognitive architecture should be designed so that, when two or more alternative inference systems are activated in a competing manner by the same input, the more specialized system will tend to override the more general one, dominating performance for that input (the *principle of pre-emptive specificity*). Therefore, when experiments are designed so that logically mediated relevance effects are placed into conflict with social contract or hazard management effects, the outputs of these content-specialized systems are predicted to override logically driven relevance effects (if they exist).
4. Both new and prior results support this prediction. Indeed, the new results show that the experimental manipulation of factors central to social contract theory and hazard management theory regulate performance on problems involving those contents, whereas the manipulation of elements central to relevance theory's account of Wason performance does not.

5. To address these new results, relevance theory (as applied to the Wason selection task) would either have to invoke content-specific inferential machinery, such as social contract and hazard management algorithms, as part of its account, or else introduce so many ad hoc assumptions as to render itself unfalsifiable.
6. These cases show that the Wason selection task does indeed illuminate the existence and properties of content-specialized algorithms, and other features of reasoning as well.

Our own view is that relevance theory is interesting and valuable, so we would not like to see it abandoned as unfalsifiable. What we will argue is that SCG are correct in pointing out that one should control for pragmatic influences and assumptions on the selection task, but that this is no different from controlling for any other experimental variable.[3] Once done, the selection task continues to be one of the most interesting and best controlled experimental paradigms for glimpsing the structure of human reasoning mechanisms.

Before proceeding, it is important to distinguish relevance theory as a general framework for understanding the phenomena of discourse comprehension from its more narrow application to interpreting performance on the Wason selection task. For clarity, we will call this application to the selection task, together with its intricate elaborations, *relevance theory*$_{Wason}$ (or RT_{Wason}), to distinguish it from relevance theory proper. Because of the elaborations, ancillary claims, and added chains of reasoning, it is possible for RT_{Wason} to be incorrect as an explanation for some or all classes of selection task performance without this either falsifying or creating difficulties for relevance theory as a content-general analysis of communication. Indeed, we will end by discussing how relevance theory (or $RT_{General}$) and content-specific theories such as social contract theory are highly compatible.

1.2. Relevance theory and the selection task

Sperber (1997, 2000) conceptualizes relevance mechanisms as evolved domain-specific mechanisms, indeed as a subunit of the theory of mind mechanism (sensu Baron-Cohen, 1995 or Leslie, 1987). This subunit is specialized for the task of discourse comprehension, and its function is to infer the speaker's communicative intentions (Sperber, 1997, 2000). Although Sperber views relevance mechanisms as specific to the domain of comprehension of communication, he also posits that they accept as input any content whatsoever as long as it is in the form of a communication from an agent. In this sense, they are *content-general*. For this reason, we will refer to the proposal that relevance theory alone (i.e. unassisted by social contract algorithms or other content-specialized inferential machinery) can explain performance on the Wason selection task as a *content-independent* or *content-general* theory. For brevity, we will refer to theories that invoke factors specific to limited content domains (exchanges, hazards, entitlements, etc.) as domain-specific

[3] An experimental design has, in effect, controlled for such factors whenever the theory being tested makes different predictions than relevance theory. In such a case, outcomes supporting the hypothesis cannot be attributed to relevance effects, which must be either absent or overridden.

theories, although content-specificity is not the only type of domain-specificity exhibited by the human cognitive architecture.

Sperber and Wilson (1986) define relevance as follows. A person judges a piece of information to be *relevant* when bringing it together with her background assumptions causes *cognitive effects*, i.e. causes her to make new inferences, such that new beliefs are adopted or prior ones abandoned. But achieving cognitive effects involves processing (*cognitive effort*), which is costly. So, all else equal, the greater the cognitive effort required to produce cognitive effects, the less relevant a piece of information will be judged to be. In interpreting utterances, the mind is designed to maximize relevance. This causes people to settle on interpretations that provide contextually sufficient cognitive effects with an economy of effort (Sperber & Wilson, 1995).

At its present stage of development, relevance theory as a cognitive theory of discourse comprehension is primarily a task analysis of what variables govern the performance of relevance mechanisms (cognitive effects, cognitive effort, background assumptions, etc.) and of the goals of processing, rather than a description of the actual computational procedures that implement and terminate the search for relevance. One exception to this is the role that logic is posited to play in comprehension (SCG; see also Sperber, 2000). In testing RT_{Wason}, we will emphasize the role of logic for two reasons: (i) it is the procedurally well-specified component of the theory; and (ii) it forms the basis for the claim that RT_{Wason} provides a general explanation of selection task performance.

According to SCG, the comprehension module is equipped with spontaneous inferential abilities (SIAs), which produce cognitive effects without deliberation or conscious reflection. SIAs that compute logical equivalences play a key role in SCG's application of relevance theory to the selection task. Logical formats and deductive rules have three interlocking design features that make them particularly appropriate for a comprehension module: they are truth-preserving, they allow certain consequences of a statement (deductive ones) to be (validly) inferred, and they can be applied to any statement (the sense in which they are content-independent).[4]

To explain performance on the Wason selection task, SCG (p. 90) "attribute to our subject a capacity to perform spontaneously specific deductive inferences involving quantifiers, and a capacity to recognize specific contradictions". This logical SIA translates conditional statements into a content-independent representational format – their logical form – and spontaneously infers their deductive consequences. "[I]n all versions of the task, subjects go beyond superficial features of the rule. They envisage testing the rule through those of its *logically derivable, directly testable consequences that they spontaneously infer*." (p. 78, emphasis added). For example, one logically derivable, directly testable consequence of *If P then Q* is that there should be no cases of *P-and-(not-Q)*. The "three consequences most pertinent to

[4] Because these design features allow one to catch contradictions, Sperber (1997, 2000) argues that a comprehension module equipped with inference procedures embodying the rules of logic would be an evolutionarily stable strategy in a world where deceptive communications are possible.

understanding subjects' performance" are listed in Table 1, Panel A (SCG, p. 78). Any of the three would cause subjects to select the *P* card; inferring consequence (b) would cause them to select the *Q* card, and inferring (c) would cause the *not-Q* card to be selected.

According to SCG, the logical SIA does not compute all three consequences in every selection task. A subject who fails to spontaneously infer consequence (c) will fail to choose the *not-Q* card; one who does spontaneously infer (b) will incorrectly choose the *Q* card. To understand card choices, one needs to understand why some deductive consequences are inferred and others are not. This is where relevance comes in: the comprehension module computes only those deductive consequences that maximize relevance.

In other words, considerations of relevance will determine how conditional rules and the selection task instructions are interpreted. SCG propose that performance on the selection task is the product of three factors: (i) the interpretation subjects assign to the task's conditional statement (i.e. which logical form is assigned); (ii) what information they deem relevant to evaluating the conditional under that interpretation; and (iii) general strategies for data selection.

(i) To settle on an interpretation, subjects "go beyond superficial features of the rule": that is, their interpretation is constrained by, but not limited to, the logical form of the conditional as stated. SCG point out that when a speaker states, 'If P then Q', there are several logically equivalent ways of interpreting this conditional, for example, (a) and (c) in Table 1. Principles of relevance will determine which interpretation is assigned. For example, many descriptive rules can be interpreted as a universally quantified statement, $\forall x\ (Px \rightarrow Qx)$ (i.e. for any x, if x has the feature *P*, then x has the feature *Q*). However, SCG (p. 55) note that "in most contexts, a general conditional is irrelevant unless it has instances."[5] In such cases, the comprehension module – which presumes that the speaker *intended* to say something it would deem relevant – would be led to interpret the statement as implying that instances of *P* exist. If it does, then 'If P then Q' would lead one to spontaneously infer (b): that there are cases of *P-and-Q*, and the conditional would be interpreted as an existentially quantified statement, $\exists x\ (Px \,\&\, Qx)$ (i.e. there exists an x, such that x has the feature *P* and x has the feature *Q*). In fact, people frequently draw the inference from a universally quantified statement to an existentially quantified statement and, though this is not a valid inference within classical predicate logic, it was considered a valid inference in scholastic logic (Rips, 1994).

(ii) The next stage in the process involves determining what information would be required to evaluate the rule. Here, problem context becomes important. Among

[5] This claim is not unproblematic. It might be more appropriate to say that a general conditional has *potential* instances. People frequently discuss counterfactual situations, and provide information about states of affairs that might occur in the future. For example, a person considering hiking in the Rockies might be told, 'If it rains, then the trails will be slippery'; a politician might say, 'If there is a nuclear war, then there will be a nuclear winter'. If a general conditional only needs to have potential (rather than actual) instances to be relevant, then relevance alone would not warrant the inference that there is at least one card with the combination of features *P* and *Q*. This would raise questions for a relevance account of performance on descriptive problems.

Table 1
Spontaneous logical inferences for descriptive and deontic rules (from SCG, p. 78, 85)

A. Descriptive rules	B. Deontic rules
(a) The rule implies, of any given card having the feature *P*, that it has the feature *Q*. (b) The rule, together with a presumption of relevance, implies in most contexts that there are cases of *P-and-Q*. (c) The rule contradicts the assumption that there are cases of *P-and-(not-Q)*.	(a″) The rule implies, of any given item having the feature *P*, that it has the feature *Q*. (b″)? The rule, together with a presumption of relevance, requires in most contexts that there be cases of *P-and-Q*. (c″) The rule forbids that there be cases of *P-and-(not-Q)*.
(a), (b), and (c) each lead to the selection of *P*; (b) leads to the selection of *Q*; (c) leads to the selection of *not-Q*, but the double negatives make it difficult to infer.	(a″) leads to the selection of *P*; (b″) is not inferred, so *Q* is not chosen; (c″) leads to the selection of *not-Q* (and *P*).

other things, subjects have to determine the intended scope of the conditional. Is it meant to apply to only the four cards provided in the task or to cards in general? If the scope of the rule is restricted to the four cards, then subjects interpreting the conditional as an existentially quantified statement $\exists x\ (Px\ \&\ Qx)$ – there is at least one card with a *P* on one side and a *Q* on the other – can falsify the rule by jointly turning over the *P* card and the *Q* card. This would allow one to disconfirm the existence of a *P-and-Q* card. However, if the scope of the conditional ranges over a larger set of cards beyond the task, then the rule interpreted as $\exists x\ (Px\ \&\ Qx)$ cannot be falsified – some card outside of the task might satisfy the rule.

(iii) To complete the task, subjects need to settle on a strategy for data selection. Should they attempt to verify the rule, or falsify it? RT_{Wason} draws attention to some of the problems of induction that often favor the search for positive instances of a claim rather than falsifying instances. While this is formally left as a free parameter, in general, RT_{Wason} suggests that subjects will settle upon that data selection strategy which is most straightforward given the information available in the problem.

1.3. Choosing the not-Q card

According to SCG (pp. 57–58), when subjects succeed in choosing the *not-Q* card, it is "by exploiting in a non-reflective, spontaneous manner a logical equivalence between quantified formulas". They are referring here to the logical equivalence given in inference (c). This is a rule of transformation, which allows one to understand that 'If P then Q' can mean 'not-[$\exists x\ (Px\ \&\ not\text{-}Qx)$]' (i.e. there are no items with the feature *P* and the feature *not-Q*; see Table 1). But when do our SIAs make this inference, and when do they not? How does RT_{Wason} explain this variability?

According to SCG, in most selection tasks employing descriptive rules people fail to spontaneously infer consequence (c) because the rule has somewhat arbitrary content, and little context. In such cases, making inference (c) has little cognitive effect (i.e. it does not allow many further inferences to be made). Moreover, it is

costly: making inference (c) usually involves processing multiple negations (high cognitive effort). In such cases, a comprehension module designed to maximize relevance would fail to infer (c), and subjects would fail to select the *not-Q* card.

By this reasoning, it should be possible to elicit *not-Q* as a response – even to a descriptive rule – by reducing the cognitive effort needed to infer (c) and by creating a pragmatic context in which knowing that cases of *P-and-(not-Q)* exist would have many cognitive effects. SCG did that by creating a discourse in which one party claims that cases of *P-and-(not-Q)* exist,[6] and a speaker asserts 'If P then Q' in order to deny that claim.

In their scenarios, a background assumption about the world is created when someone (either a character or the narrator) asserts that cases of *P-and-(not-Q)* exist. In these scenarios, the existence of such cases is either bad, shameful, and/or unusual (e.g. on the job errors, virgin mothers created by a cult leader, unemployed citizens, bachelors volunteering to care for children) – the kind of things an interested party might wish to deny. Being told that this background assumption is incorrect – i.e. that these bad, shameful, or unusual things do not exist – would therefore lead one to revise one's beliefs. That is, it would have cognitive effects: it would be *relevant*. Thus, when the central character in the story responds to the assertion that cases of *P-and-(not-Q)* exist by saying, 'If something is *P* then it is *Q*', inference (c) is triggered,[7] and the conditional is interpreted as meaning 'I *deny* your charge that there are cases of *P-and-(not-Q)*'. This is because the comprehension module is designed to consider cognitive effects in their order of accessibility (following a path of least effort), and settle on an interpretation that maximizes relevance. (Inferences (a) and (b) are passed up as less relevant: implying or stating that cases of *P-and-Q* exist is not responsive to the charge that cases of *P-and-(not-Q)* do exist, particularly compared to a flat denial.)

The cognitive effort of processing an explicit 'not' in the logical statement 'not-[∃x (*P*x & *not-Q*x)]' can be reduced by translating it into its implicit form: '*Denied*-[∃x (*P*x & *not-Q*x)]' (further reductions in cognitive effort can be made by translating *not-Q* into an implicit form, e.g. using 'virgin' in the text, rather than 'has never had sex'). This statement can be directly falsified by finding an item with the features *P* and *not-Q*, so subjects should test this statement by selecting the only cards that could potentially have those features: the *P* card and the *not-Q* card. In several

[6] Mentioning cases of *P-and-(not-Q)* in the problem's text is a potential confound in these experiments. Subjects could be choosing the *not-Q* card merely because the existence of cases of *P-and-(not-Q)* have been brought to mind by having been previously highlighted in the text (rather than because the discourse structure triggers an interpretation of the conditional as a denial). Further experiments would have to be done to eliminate this alternative.

[7] It is not clear how the mind knows which inferences will maximize cognitive effect unless it actually makes them. If one spontaneously infers (c), then one can assess its cognitive effects; but if it is not inferred, how does one know what its effects might have been? As a design principle operating over phylogenetic time, one can see how a selection pressure of this kind might have shaped what information the mind is drawn to and what inferences are habitually made in a given domain. But as an online process, it is unclear how this prediction of relevance theory could be implemented. This is one reason we believe that most relevance effects are the product of evolved, content-specific and domain-specific machinery.

experiments, SCG find that they do, as long as the conditional can be assigned the logical form of a denial. For example, 57% of their subjects answered '*P & not-Q*' in a task that employs an indicative conditional that clearly is not the sort of deontic prescription that has previously been shown to elicit this pattern of selection (SCG, Experiment 4). The selection task employed in this experiment describes a situation in which a machine is supposed to print cards with the number 6 on the front and the letter E on the back, but it is malfunctioning by failing to print an E on the back. After the repairman has fixed the machine he denies that it is continuing to malfunction by assuring: *If a card has a 6 on the front, it has an E on the back* – a conditional very similar to that originally employed by Wason (1968), which had elicited the answer, '*P & not-Q*', from less than 10% of subjects tested.

The machine problem and the other content domains SCG used fall outside of the scope of phenomena that the domain-specific deontic theories were designed to explain (e.g. Cheng & Holyoak, 1985; Cosmides, 1989; Cosmides & Tooby, 1989; Manktelow & Over, 1991). For this reason, SCG's existing experiments cannot be used to decide whether (for example) social contract theory and relevance theory each characterize separate mechanisms that produce independent effects. New experiments must be designed that could potentially measure the relative robustness of each effect when both mechanisms are predicted to be operating, or when they predict divergent outcomes.

SCG's claim that the selection task cannot reveal the presence of content-specialized inferential machinery has merit only if it is true that content-free logical SIAs are sufficient to explain all content effects. We, in contrast, think it can be shown that the content-free SIAs invoked by SCG are only sufficient to explain some content effects, at best. Indeed, we suspect that their use of discourses involving assertions and denials may be one of the only ways of triggering inference (c) that is truly content-general. This is because the situation implies a disagreement and the assertion explicitly states exactly *which* background assumption out of potentially a very large set the main character is challenging by stating the conditional, leading to the interpretation that the conditional was meant to deny that assertion. When this conversational device is not used, there is no content-general way for listeners to logically isolate which background assumption is at issue and which interpretations are intended.

These are exactly the gaps that content-specific theories fill. Specialized inferential systems were proposed exactly because they are capable of solving an array of computational problems that plague content-independent systems (such as those dependent on logic), including the combinatorial explosion of possible computations and the difficulty of isolating those inferential pathways that are relevant to the problem at hand. They do so by incorporating computational machinery that is *ecologically rational*.

1.4. Ecological rationality and two applications: social contract theory and hazard management theory

RT_{Wason}'s explanation of the selection task achieves its content-generality by

invoking traditional concepts of rationality, such as conformity to the canons of logic. Over the last decade, however, a non-traditional view of rationality – *ecological rationality* – has been developing, and with it a different approach to understanding performance on the selection task (Barkow et al., 1992; Cosmides & Tooby, 1996a,b; Gigerenzer & Hoffrage, 1995; Gigerenzer, Hoffrage & Kleinbolting, 1991; Gigerenzer, Todd & the ABC Research Group, 1999; Sperber, 1994; Tooby & Cosmides, 1992, 2000). According to this view, the human cognitive architecture is densely populated with a large number of evolved, content-specific, domain-specific inference engines (or evolved mechanisms for their developmental acquisition), in addition to whatever more domain- or content-general inferential competences may exist. Each of these is designed to operate over a different class of content. For example, converging evidence from cognitive neuroscience, cognitive development, and evolutionary psychology suggests that there are separate inference systems for reasoning about objects (Baillargeon, 1986; Spelke, 1990), physical causality (Brown, 1990; Leslie, 1994; Leslie & Keeble, 1987; Talmy, 1988), number (Gallistel & Gelman, 1992; Wynn, 1992, 1995), the biological world (Atran, 1990; Gutheil et al., 1998; Hatano & Inagaki, 1994; Keil, 1994; Springer, 1992), the beliefs and motivations of other individuals (Baron-Cohen, 1995; Gergely, Nadasdy, Csibra & Biro, 1995; Leslie, 1987), hazard management (Cosmides & Tooby, 1997; Fiddick, 1998), and social interactions (Cosmides & Tooby, 1989, 1992; Fiske, 1991; for review, see Hirschfeld & Gelman, 1994). Each is a 'spontaneous inferential ability' (SIA), in SCG's sense, but they are not *logical* SIAs. They create the background assumptions on which (for example) computations of relevance depend, and most of them embody rules of transformation that are not licensed by the predicate calculus, but that are nonetheless appropriate to their given content domain.

When activated by content from the appropriate domain, these inference engines impose special and privileged representations during the process of situation interpretation, define specialized goals for reasoning tailored to their domain, and make available specialized inferential procedures that allow certain computations to proceed automatically or 'intuitively' and with enhanced efficiency over what a more general reasoning process could achieve given the same input (Cosmides, 1985, 1989; Cosmides & Tooby, 1989, 1992; Gigerenzer & Hug, 1992; Tooby & Cosmides, 1992). While the designs of these systems may not embody content-independent norms of rationality, such as the predicate calculus or Bayes's rule, they are *ecologically rational* (Cosmides & Tooby, 1996b; Gigerenzer et al., 1999; Tooby & Cosmides, 2000). That is, each embodies functionally specialized design features that reflect the task demands of the adaptive problem it evolved to solve, including assumptions about the evolutionarily long-term ecological structure of the world. As a result, when operating within the domain for which they evolved, they solve adaptive problems reliably, efficiently, and with limited information.

Two of these proposed inference engines – one for reasoning about social exchange, the other for reasoning about hazards – have been applied to understanding performance on the Wason selection task. Social contract theory (Cosmides & Tooby, 1989) and hazard management theory (Cosmides & Tooby, 1997; Fiddick,

1998) are accounts of the domain-specific representational formats and inference procedures activated by these content domains. They are intended to explain the intuitions that people have when interpreting these situations and reasoning about them.

1.4.1. Social contract theory

Although zoologically rare, social exchange – cooperation for mutual benefit – is an ancient and pervasive feature of human social life. It has been proposed that the human computational architecture contains an expert system designed for reasoning about social exchange – the *social contract algorithms* – one component of which is a subroutine specialized for cheater detection (Cosmides, 1985, 1989; Cosmides & Tooby, 1989, 1992, 1997). Social contract theory is a computational theory specifying what design features an expert system functionally specialized for engaging in social exchange should have (see Cosmides, 1985; Cosmides & Tooby, 1989).

In their computational theory, Cosmides and Tooby (Cosmides, 1985; Cosmides & Tooby, 1989) defined a social exchange as a situation in which, in order to be entitled to receive a *benefit* from another individual or group, an individual is obligated to satisfy a requirement of some kind (often, but not necessarily, at some cost to him- or herself). Those who are rationing access to the benefit impose the requirement because its satisfaction creates a situation that benefits them.[8] A *social contract* expresses this intercontingency, and has the form, 'If you accept the benefit, then you must satisfy the requirement' or, equivalently, 'If you satisfy the requirement, then you are entitled to the benefit'. These statements are not *logically* equivalent. Rather, there is a set of inference rules *specific to this domain* that license each as a translation of the other. Some of these rules of transformation are specified in Table 2 (for sources, see Cosmides, 1985; Cosmides & Tooby, 1989). These play a key role in the experiments reported herein.

Social contracts can be explicitly made (as when two individuals agree to trade), or understood implicitly (as when one feels obligated to return a kindness when the opportunity arises). Many superficially different forms of social interaction fit this general structure, e.g. economic trades, reciprocal gift-giving, helping with the expectation that the favor will someday be returned, and certain social laws (situations in which a social group controls access to a benefit, and restricts access to it to those who have satisfied some requirement) (see Appendix A).

Social contract theory is based on the hypothesis that the human mind was designed by evolution to reliably develop a cognitive adaptation specialized for reasoning about social contracts. To demonstrate the existence of an adaptation one needs to provide evidence of *special design* (Dawkins, 1986; Williams, 1966). It is an engineering standard: special design is evidenced by a set of features of the phenotype that combine to solve an element of a specific, adaptive problem particularly well, and in a way highly unlikely to have arisen by chance alone. In the case of social exchange, evolutionary analyses indicate that one very important

[8] The role of costs in social contract theory has sometimes been misunderstood (e.g. Cheng & Holyoak, 1989) (see Appendix A).

Table 2
Exchanges: inferences licensed by social contract algorithms

'If you give me P then I will give you Q' (= 'If I give you Q then you give me P')[a]
Either expression means (entails) the following:
1. I want you to give me P,
2. My offer fulfills the cost/benefit requirements of a sincere contract *(listed in Table 3)*
3. I realize, and I intend that you realize, that 4–9 are entailed if, and only if, you accept my offer:
4. If you give me P, then I will give you Q,
5. By virtue of my adhering to the conditions of this contract, my belief that you have given (or will give) me P will be the cause of my giving you Q,
6. If you do not give me P, I will not give you Q,
7. By virtue of my adhering to the conditions of this contract, my belief that you have not given (or will not give) me P will be the cause of my not giving you Q,
8. If you accept Q from me, then you are obligated to give me P (alternatively: If you accept Q from me then I am entitled to receive P from you),
9. If you give me P, then I am obligated to give you Q (alternatively: If you give me P then you are entitled to receive Q from me).

What does it mean for you to be *obligated* to do P?
a. You have agreed to do P for me under certain contractual conditions (such as 1–9), and
b. Those conditions have been met, and
c. By virtue of your not thereupon doing P, you agree that if I use some means of getting P (or its equivalent) from you that does not involve getting your voluntary consent, then I will suffer no reprisal from you. *(OR: By virtue of your not thereupon giving me P, you agree that if I lower your utility by some (optimal) amount X (where* $X > Byou$ *– your unearned gains), then I will suffer no reprisal from you.)*

What does it mean for you to be *entitled* to Q?
d. I have agreed to give you Q under certain contractual conditions (such as 1–9), and
e. Those conditions have not been met, and
f. By virtue of my not thereupon giving you Q, I agree that if you use some means of getting Q (or its equivalent) from me that does not involve getting my voluntary consent, then you will suffer no reprisal from me *(OR: By virtue of my not thereupon giving you Q, I agree that if you lower my utility by some (optimal) amount X (where* $X > Bme$ *– my unearned spoils), then you will suffer no reprisal from me.)*

[a] 'Give' takes three arguments: two agents and the entity given. From this perspective, it is important to preserve the correct binding of agents to items of exchange, and the intercontingent nature of the giving. It is *not* relevant which agent is the subject in the if-clause and which in the then-clause. Furthermore, the entailments all hold, regardless of who fulfills their part of the contract first (i.e. tense is irrelevant, unless it is specified that order falls under the terms contract).

problem to be solved is the detection of cheaters: individuals who accept benefits without satisfying the requirement that their provision was made contingent upon (e.g. Axelrod, 1984; Axelrod & Hamilton, 1981; Boyd, 1988; Tooby & Cosmides, 1996; Trivers, 1971; Williams, 1966). Therefore, one prediction of social contract theory is that people should show a special ability to detect cheaters in a social exchange – people should possess a 'look for cheaters' algorithm. There are other predictions as well: by virtue of having social contract algorithms, people should readily or automatically infer the many implications of an exchange (Table 2), they

should be able to understand the benefits and costs to each party (Table 3), they should wish to avoid and/or punish cheaters, and so on (for a more complete list, see Cosmides, 1985; Cosmides & Tooby, 1989).

The Wason selection task has been used to test for the existence of the hypothesized cheater detection algorithm. In the selection task, a social exchange can be expressed by a conditional rule, as described above. Subjects encountering such a rule, in a scenario where cheating is possible, should see that there are two potential cheaters among the four individuals represented by the cards: the individual who has accepted the benefit and the individual who has not satisfied the requirement.

An algorithm designed to look for cheaters should select the *benefit accepted* card and the *requirement not satisfied* card, *regardless of which logical category each happens to fall into*. If the conditional rule reads 'If you accept the benefit, then you must satisfy the requirement', then these cards will correspond to the logical categories *P* and *not-Q*, and the subject will appear to have reasoned logically. If it reads 'If you satisfy the requirement, then you are entitled to the benefit', then these cards will correspond to the logical categories *Q* and *not-P*, and the subject will appear to have reasoned illogically (see Fig. 1). But in either case, the choice of these cards is the adaptively correct response: these, and only these cards represent potential cheaters. These predictions have now been verified in a number of studies (e.g. Cosmides, 1989; Cosmides & Tooby, 1992; Gigerenzer & Hug, 1992; Platt & Griggs, 1993).

Note that the definition of cheating *does not map onto the logical definition of violation* (the latter being a true antecedent paired with a false consequent). Cheating is a content-dependent concept: there must be an illicitly taken *benefit*. This, and only this, counts as cheating. Logical categories and definitions of violation form an orthogonal representational dimension.

By providing an assay of when the cheater detection routine is activated and inactivated, the selection task has also allowed researchers to test many other hypotheses about the features of social contract algorithms (e.g. those specified in Tables 2 and 3). It is a curiosity that those who have commented on social contract theory have focussed almost exclusively on the cheater detection procedure, without any mention of the well-specified theory of interpretation that the theory has, since its inception, provided (see Cosmides, 1985; Cosmides & Tooby, 1989). In the experiments reported herein, we demonstrate how both components of the theory – the interpretive inferences and the post-interpretive cheater detection mechanism – are useful in predicting and understanding selection task performance.

1.4.2. Hazard management (precaution) theory

We also believe there is an evolved inferential system specialized for reasoning about hazards, and the precautions one can take to minimize the risk of being exposed to hazards (Cosmides & Tooby, 1997; Fiddick, 1998). This is hypothesized to be a separate system from the social contract system, with its own distinct architecture, representational format, and licensed inferential procedures. Indeed, reasoning about what we call *precaution rules* also elicits highly organized subject

Table 3
Sincere social contracts: cost/benefit relations for an exchange of goods when one party is sincere, and that party believes the other party is also sincere[a]

My offer:	*'If you give me X then I'll give you Y'*			
	Sincere offer		Sincere acceptance	
	I believe:		*You believe*:	
P	B_{me}	C_{you}	B_{me}	C_{you}
not-P	0_{me}	0_{you}	0_{me}	0_{you}
Q	C_{me}	B_{you}	C_{me}	B_{you}
not-Q	0_{me}	0_{you}	0_{me}	0_{you}
Profit margin:	*Positive*: $B_{me} > C_{me}$	*Positive*: $B_{you} > C_{you}$	*Positive*: $B_{me} > C_{me}$	*Positive*: $B_{you} > C_{you}$
Translation:				
My terms	'If B_{me} then C_{me}'		'If B_{me} then C_{me}'	
Your terms	'If C_{you} then B_{you}'		'If C_{you} then B_{you}'	

[a] Costs and benefits are relative to a baseline that each party believes would pertain in the absence of an exchange (the zero-level utility baseline). B_x = benefit to individual x; C_x = cost to individual x; 0_x = no change from x's utility at baseline. A contract has been *sincerely* offered and accepted when both parties are being truthful about their baselines and when each believes the B > C constraint holds for the other (Cosmides, 1985; Cosmides & Tooby, 1989).

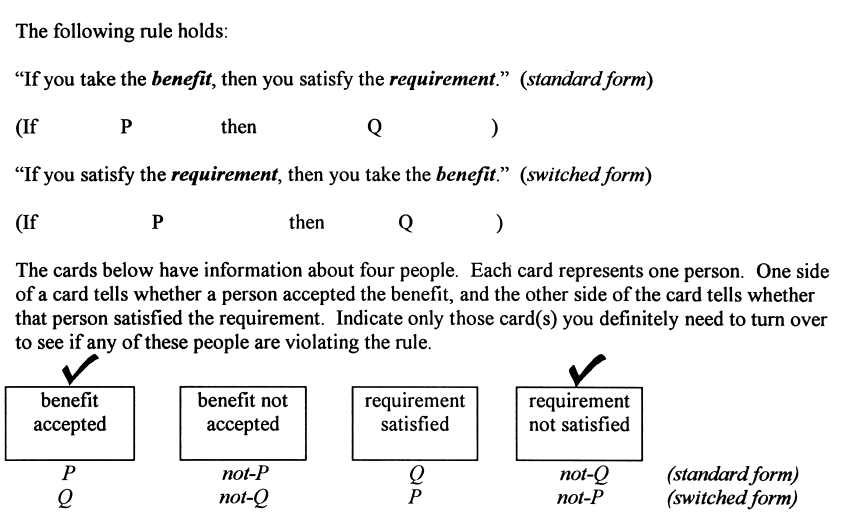

Fig. 1. Abstract structure of a selection task using a social contract rule. For example, assuming 'you' are the potential cheater, 'If I give you $10, then you give me your watch' would be standard, and 'If you give me your watch, then I'll give you $10' would be switched. To detect cheaters, one needs to choose the *benefit accepted* card and *the requirement not satisfied* card, regardless of their logical category. Choosing these cards results in a logically correct answer, '*P & not-Q*', for a standard social contract. But for a switched social contract, the same choice results in a logically incorrect (but adaptively sound) answer, '*not-P & Q*'.

performance, and yet a precaution rule is clearly not a social contract rule (Cosmides & Tooby, 1997; Fiddick, 1998).

As we analyze it, a precaution rule is of the form: *If a valued entity is subjected to a hazard, then taking an appropriate precaution lowers the risk of harm.* When given a rule of this form in a selection task, subjects reliably select the *subject to hazard* and *did not take precaution* cards, which correspond to the logical categories *P* and *not-Q*, respectively (for rules we would categorize as precautions in this sense, see Cheng & Holyoak, 1989; Manktelow & Over, 1988, 1990; Stone et al., 2000).

Because there are conditions in which social contract rules and precaution rules elicit the same logical performance on the selection task, it is tempting to hypothesize that one psychological mechanism underlies reasoning in both content domains. Cheng and Holyoak (1989), for example, claim that rules that we would categorize as social contracts and rules that we would categorize as precautions are simply instances of what they hypothesize to be a single, psychologically real, inclusive category: permission rules.[9] Similarly, Manktelow and Over (1990, 1991) propose that the more inclusive class being operated over is any deontic rule with utilities. Empirically, social contract rules and precaution rules both elicit strong content effects.

[9] Cheng and Holyoak (1985) also propose obligation schemas. In our own past experiments on precaution rules, we have been careful to phrase them such that they would be categorized according to their theory as permissions (e.g. Stone et al., 2000).

An alternative, adaptationist account of these empirical results posits two independent domain-specific inference systems: one specialized for reasoning about social contracts, and another specialized for reasoning about hazardous situations (Cosmides & Tooby, 1997; Fiddick, 1998). Because each problem poses different computational requirements, the features of these two reasoning systems should be different. The elements of a cognitive adaptation for reasoning about hazards should use representational primitives and specialized inferential procedures that derive from the evolutionarily enduring ecological structure of managing dangers. Hazard management was an important adaptive problem during human evolutionary history. Everyday tasks necessary for survival and reproduction – foraging, hunting, attracting mates, protecting one's children against predators and aggressors, and myriad others – exposed our ancestors to a variety of hazards. Cognitive architectures that included inferential methods for cost-effectively minimizing these hazards would certainly have had a selective advantage. A large proportion of naturally recurrent dangers could be reduced – and sometimes nullified – by taking appropriate precautions. But for risk to be reduced, the precaution has to be taken. One might therefore expect the evolution of various checking routines, designed to reduce the relevant risk – whether it is to oneself, other individuals, or valuable resources (tools, food caches, etc).

Note that the core terms of hazard management theory (danger, precaution in effect, etc.) are different than the core terms of social contract theory (e.g. benefit, requirement, cheating, etc.). This allows a series of principled predictions to be derived about contrasting performance on social exchanges and precaution rules. For example, in social exchange, any item, action, or state can count as a benefit or cost to the interactants because values are in the eye of the holder. But facts about the world – and not the desires of agents in the scenario – are relevant to hazard management routines, because their purpose is danger reduction. This suggests that their checking routine will be more difficult to activate – and performance will be lower – when the 'precaution' is not judged by the *subject* as effective against the hazard (regardless of what agents in the scenario might think). Indeed, Rutherford, Cosmides and Tooby (1996, 2000) have found evidence confirming this prediction. In this and many other ways, the computational requirements for detecting violations of precaution rules are different from those for cheater detection, so different, domain-specific mechanisms should have evolved to handle each (Cosmides & Tooby, 1997; Fiddick, 1998).

As we will discuss in the next section, the view that there are two distinct, evolved expert systems operating over these two content domains – one for hazard management and a different one for social exchange – predicts and accounts for a series of findings that single-mechanism theories do not. It also plays a central role in the logic of the experiments presented herein.

1.5. Strategies for evaluating competing theories

Social contract theory is a content-specific theory: it was developed to predict and explain reasoning about social contracts, and only social contracts. Because social

contract algorithms are not designed to be activated outside of the domain of social exchange, social contract theory was not intended as an explanation of content effects or of apparently logical performance in other domains. Although social contract theory provided a parsimonious explanation for a series of newly predicted content effects, and indeed for all of the Wason content effects that had been detected at the time of the theory's introduction in 1983, we nevertheless fully expected that other content effects would be subsequently discovered or intentionally generated for other evolutionarily significant domains. Social contract theory, and its associated experiments, were offered as an existence proof and an initial, illustrative application of a more general, evolutionary-functional approach to reasoning, in the hope that others would join in the effort to construct domain-specific theories appropriate to other evolutionarily important content domains. We ourselves, together with our colleagues, have been working on additional evolutionarily significant domains, such as hazards and threats, and on adaptationist grounds expect there to be many others. Moreover, the social contract effect is a robust one. The evaluative strategy of attempting to falsify social contract theory through identifying social contract contents that do not elicit the effect has consistently failed, and despite many attempts at falsification, the social contract effect has been replicated scores of times, using different rules, by different researchers on different populations (e.g. see Sugiyama et al., 2000 for tests on a non-literate hunter-horticulturalist population and Stone et al., 2000 for a report of a normed set of 18 social contract rules, 23 precaution rules, and 24 abstract/descriptive rules).

Despite the fact that social contract theory was developed to predict and explain performance in a single domain, some, to our surprise, have argued that it is a failing or even a falsification of the evolutionary approach to reasoning that social contract theory does not account for performance on rules with other types of content (e.g. Cheng & Holyoak, 1989; Kirby, 1994; Manktelow & Over, 1990, 1991; Pollard, 1990; Rips, 1994). Obviously, the discovery of some content effects outside of the domain of social exchange does not falsify social contract theory, since such discoveries are what those who hold an adaptationist approach to reasoning expected in the first place. Such discoveries do mean, however, that there are two major competing theoretical possibilities to be examined. The first possibility is that there are, as we have predicted, evolved inferential specializations for other adaptively significant domains (in addition to those for social exchange), and that each of these generates its own differently organized but equally principled content effects.

A second possibility is that social contract effects, despite their consistency with prediction, are not caused by social contract algorithms, but are instead generated by some more general process that accounts for a wider, more inclusive set of content effects that simply happens to include social exchanges – along with many other contents – as one incidental subset of no special psychological significance. This is the view taken by advocates of permission schema theory, RT_{Wason}, mental models theory, explanations invoking utilities and deontic logic, and so on. After all, one explanation for an inclusive set of cases seems more parsimonious than an explanation involving a series of independent cognitive specializations, each of which accounts for a more restricted set of cases. More fundamentally, this approach is

preferred by many cognitive scientists because there is widespread resistance to the proposal that the evolved architecture of the human mind contains anything so content-specific as social contract algorithms.

Nevertheless, given that a single mechanism account and a multiple mechanism account both predict the existence of some content effects in domains other than social exchange, the mere existence of such effects cannot decide between theories, as many would have them do (e.g. Cheng & Holyoak, 1989). So, what forms of evidence could be used to decide among proposals?[10]

1.5.1. Theories differ in their scope of application: the case of the missing content effects

The single and multiple mechanism proposals differ in their account of the underlying representations and, therefore, in their scope of application. For example, the scope of a permission schema is much larger than the scope of social contract or precaution representations. If the permission schema representation (proposed by Cheng & Holyoak, 1985) is psychologically real, then rules that are perfectly good permission rules should reliably elicit high levels of performance, even when they lack the additional features of a social contract or a precaution. Yet they do not. Indeed, it is easy to construct ordinary permission (and obligation) rules according to the theory of Cheng and Holyoak (1985) which nevertheless fail to elicit the content effects that they predict should occur (Cosmides, 1989; Cosmides & Tooby, 1992; Gigerenzer & Hug, 1992). The finding that a theory's predictions fail over a large part of the range where its core theoretical elements say it should apply is an especially strong falsification.

To generate large sets of non-performing permission and obligation rules, one only needs to generate normal deontic rules that cannot be given a social contract or precautionary interpretation by the subject.[11] In many cases, this can be done by manipulating a single factor that is crucial to representing a rule as a social contract – e.g. the presence of a benefit – but irrelevant to representing it as a permission (or obligation) rule (Cosmides & Tooby, 1992). The far more restricted sets of social exchange and precautionary rules, in contrast, appear always to produce the predicted content effect throughout the domains defined by their theoretical scope. The explanation for heightened performance on social contracts offered by Cheng and Holyoak (1989) – that they elicit high levels of performance because they are simply instances of permission rules – cannot be correct, because permission rules that are not social exchanges or precautions do not elicit the predicted effect.

Permission rules are deontic. Therefore, the same finding – that people are poor at detecting violations of deontic rules that are neither social contracts nor precautions

[10] Although important, we will not be discussing the developmental and cross-cultural methods of theory evaluation in this article, because the research carried out so far using these approaches cannot be used to evaluate relevance theory.

[11] We expect that there may be other adaptive domains, yet to be identified, outside of social contracts and precautions. The point is that the category of permissions and obligations is very large compared to the adaptively significant domains, so one can find indefinitely many rules that permission schema theory predicts should work, that do not.

– creates difficulties for other deontic reasoning theories as well. This includes RT_{Wason}. As we will see, SCG's analysis implies that the same logical form will be assigned to all deontic rules (especially when *P* denotes a widespread activity). To exclude all deontic rules *other than* social contracts and precautions, they would have to introduce a host of ad hoc assumptions. Alternatively, SCG could alter RT_{Wason}, making it incorporate social contract and hazard management algorithms as active theoretical components – but then their account would rely on content-specialized inferential machinery rather than being independent of it.

1.5.2. Dissociations and selective deficits

The hypothesis that there are two systems, one specialized for hazard management, and one for social contracts, leads immediately and straightforwardly to the prediction that dissociations between the two can be generated either by brain damage or by experimental manipulations. In contrast, if there is only a single cognitive system (such as a facility with deontic logic, information gain, or permission rules), and that system operates on rules or problems by virtue of their being members of some more inclusive class, then dissociations that appear along the fault lines predicted in advance by the domain-specific theories are difficult to account for, *even in a post hoc manner*.

For example, if a single, more inclusive mechanism is governing performance, then cognitive impairment should reduce performance on all members of the inclusive class, not just on social contracts. Yet Stone and colleagues have identified a patient with focal brain damage whose ability to detect violations of precaution rules is normal, but whose ability to detect cheaters on (logically equivalent) social contracts is sharply impaired (Stone et al., 1996, 2000). Similarly, by using a priming paradigm and the selection task, Fiddick et al. (1995, 2000) were able to create the functional equivalent of a neurally based double dissociation in reasoning about social contracts and precautions (see also Fiddick, 1998). Results like these support the view that different sets of reasoning procedures are applied to the two classes of rules, even when the rules are assigned the same logical form.

Such results indicate that the human mind treats social contracts and precautions as distinct classes, even when researchers do not. A researcher's categorization scheme may seem reasonable, yet not turn out to be psychologically real. For example, Manktelow and Over (1990) categorize precautions (and social contracts) as instances of the general class of deontic rules. But, as we will see in Experiment 2, it is quite possible to remove the 'deonticity' from a precaution rule without changing the pattern of content effects. This supports the view that it is not the (supposed) deontic nature of certain rules that elicits the effect, but rather their conformity to the structure of precautions or social exchanges.[12]

[12] In this article, we use the term 'deontic' to refer to a set of rules in various experiments that others have interpreted as involving entitlement and obligation, without accepting that this categorization is psychologically real. Such categories are built up out of domain-specific machinery whose nature derives from its evolved function. For this reason, social contracts and precautions have different properties, and do not map onto concepts in deontic logic in any simple or uniform manner.

1.5.3. Manipulations that regulate content effects

A theory gains credibility to the extent that experiments show that the manipulation of variables relevant to that theory, but irrelevant to others, cause predicted changes in the dependent variable. Credibility is enhanced or reduced according to the degree that the predicted effects are precise, reliable, and unique to that theory.

A number of different elements – the presence of a benefit in the rule, whether the benefit was taken intentionally or as an innocent mistake, the perspective of the rule-interpreter, and so on – are central to social contract theory but irrelevant to competing theories. It has been repeatedly shown that manipulating these variables can trigger or eliminate content effects according to patterns predicted in advance by social contract theory (Cosmides, 1989; Cosmides & Tooby, 1992, 1997; Fiddick, 1998; Gigerenzer & Hug, 1992).

Providing instances that confirm the predictions of a particular theory – as we have done for social contract theory and as SCG have done for RT_{Wason} – clearly adds credibility to that theory. But it does not rule out the possibility that both theories are correct. To illustrate, consider the hypothesis that monkeys have an evolved fear response that is functionally specialized for snakes. The discovery that heart rates in laboratory-reared monkeys accelerate in response to their first exposure to snakes supports this hypothesis, and is not in the least undermined by the discovery that their heart rates also accelerate in response to sudden loud noises. Nor do both discoveries jointly require a single, more overarching general explanation in which snakes and sudden noises play no specific role.

Similarly, multiple domain-specific inference mechanisms – relevance procedures in the comprehension module, social contract procedures, and so on – may well coexist side by side as independent domain-specific components of the same human cognitive architecture. For this reason, positive evidence for one is not necessarily negative evidence for another. The two sets of theories may compete as intersecting explanations for some sets of observations, but the truth of one does not entail the falsity of the other. In short, some sets of theories are mutually compatible, and social contract theory, hazard management theory, relevance theory and even RT_{Wason} could all be simultaneously true.

How then do we decide whether both are true, both false, or one is true while the other is false? To answer this question, we will need to derive one additional prediction from the ecological rationality approach.

1.5.4. Predictions about conflicting mechanisms: the principle of pre-emptive specificity

A domain is defined as an identifiable set of phenomena or situations that share a certain evolutionarily enduring ecological structure or statistical composite of characteristics in common (Cosmides & Tooby, 1997; Sperber, 1994; Tooby & Cosmides, 1992). Consider, for example, the set of *social exchanges*, the set of *utterances*, and the set of *physical objects*. From a computational point of view, each of these sets has a common structure of properties shared by its members, but different from the shared characteristics that typify members of the other sets. *Utterances*, for example, are produced by human beings – not by rocks – and reflect

the intentions of the speaker. This bundle of common properties makes the domain-specific theory of mind mechanism useful in interpreting utterances, but useless for inferring the behavior of rocks.

Evolved cognitive specializations get their enhanced problem-solving power by being able to exploit those relationships that characterize a domain but may be limited to it. They accomplish this through applying computational short-cuts (reduced problem spaces, tailored representational formats, conceptual primitives, specialized inferential procedures) that work well within the domain. These short-cuts do not necessarily work outside their target domain, because instances from outside the domain do not necessarily manifest the evolutionarily enduring features that the short-cuts evolved to depend upon. Thus, a language acquisition device can assume that language inputs share the invisible but evolutionarily enduring properties of universal grammar; an object mechanics module can assume that two solid objects cannot pass through each other without deformation; and social exchange algorithms embody assumptions that in social exchanges, each party is seeking a new, more beneficial state of affairs for herself than what she could have expected in its absence.

Because the mind is populated by many computational systems in a heterarchical array, a problem may fit the input conditions for many different cognitive mechanisms at once. For example, an input may have content marking it as a social exchange, activating social contract algorithms, while at the same time it is marked as an utterance, activating discourse comprehension systems and logical equivalences.

All else equal, the more characteristics the members of a set have in common, the more powerful a problem-solver specialized to exploit those characteristics can be, and so the richer the inferences that can be drawn about individual members of the set from limited inputs when that specialization is applied. Significantly, however, there is a trade-off between the number of characteristics held in common and the breadth of the set: the narrower the set, the more its members will tend to have in common (e.g. instances of the narrower set, [social exchanges], share more in common than instances of the broader set, [social exchanges, utterances, and objects]). As a result, evolved problem solvers specialized for social exchanges can deliver richer inferences than ones whose inference procedures are limited to those that apply validly across a broader set.

While the issue of which components become activated, and how they interact is a complex one, there is a rule of thumb that emerges from an ecological rationality perspective.

With certain qualifications, it follows that:

1. The most informative procedures that can be applied to an input will tend to be those that treat it as a member of the narrowest domain for which there exists a domain-specific inference engine designed to solve its inferential problems.
2. In the absence of something more specialized, the system falls back on weaker inference systems that apply over a broader class of problems that have fewer characteristics in common.

3. Therefore, a well-engineered problem-solving system should deploy, to the extent possible, the most specialized problem-solving machinery that is activated by the problem at hand, because on average, it will be more knowledgeable than the alternative, more general problem-solvers that also apply (for discussion, see Tooby & Cosmides, 1992, 2000).[13]

This principle of cognitive design – what we will call the principle of *pre-emptive specificity* – should be expressed in design features throughout the cognitive architecture. It applies to the problems herein. Social contract algorithms and hazard management algorithms are more content-specialized than relevance mechanisms, whose domain is all content that arrives via communication from an agent. What this means is that, for problems in which social contract algorithms or hazard management algorithms are set in opposition to relevance effects, *the social contract or hazard algorithms are predicted to pre-empt or override the relevance mechanisms' logical SIAs, if they exist*.

There might be cases in which the evidence for social contract theory or hazard management theory is confounded with the conversational pragmatics of the selection tasks used to test the theory, just as SCG claim (p. 88). However, the confound works both ways: content-specific cognitive adaptations may be doing the work attributed to conversational pragmatics on deontic versions of the selection task. Unconfounded experiments are needed to demonstrate which effect dominates.

So, in the design of the experiments that follow, our goal was to pit the hypothesis that social contract algorithms and hazard management algorithms strongly regulate reasoning on the Wason selection task against the counter-hypothesis that RT_{Wason} explains all content effects on the Wason selection task (including those that have been attributed to the action of social contract and hazard management algorithms). The RT_{Wason} counter-hypothesis, if true, would remove one source of evidence otherwise believed to support the hypothesis that social contract algorithms exist. The evaluative strategies that we will pursue are (i) remove factors central to the RT_{Wason} account while preserving the social exchange and precaution structures of problems, to see if the predicted content effects remain, and (ii) construct (and review) reasoning problems where RT_{Wason} makes predictions that contrast with social contract theory and with hazard management theory, and then see which set of predictions are confirmed.

If our predictions are confirmed, these experiments would support the view that social contract and hazard management algorithms exist; they would also falsify the counter-hypothesis that content effects on social contract rules and precaution rules

[13] Obviously, this is a complex engineering issue in which other principles are also expected to regulate which inference systems are activated (Tooby & Cosmides, 2000), and domains cross-cut each other in ways that cannot always be ordered as narrower or broader. Moreover, many processes interact or operate in parallel on inputs, as we expect discourse comprehension mechanisms and social contract algorithms to do on reasoning problems. Nevertheless, even though outputs may be received from many parallel systems, jointly influencing the outcome, the richest outputs will usually be from the content-specific inferential systems, and so they should play a major and experimentally detectable role in subsequent information integration.

can be explained by invoking nothing other than content-general relevance factors. However, it is important to bear in mind that these experiments would not falsify relevance theory itself – either in general, or the more specific idea that relevance mechanisms might explain some content effects on the selection task *outside of the domains of social contracts and precautions*.

Here is why: if relevance mechanisms and social contract algorithms (for example) both exist, the principle of pre-emptive specificity predicts that social contract effects, being stronger, will override the effects of relevance mechanisms on social contracts. But experimental results in conformity with this prediction (or the parallel one about precautions) cannot speak to the question of whether relevance mechanisms are nevertheless real, and operate detectably on other contents outside of domains in which they are pre-empted by evolved specializations. For example, we find the RT_{Wason} account of subject performance on most descriptive rules plausible, and worth exploring.

1.6. Relevant with respect to which representations? Areas of controversy, agreement, and complementarity

In assessing the relative merits of the content-general and content-specific proposals, it is crucial that one is clear about what is really at issue. Interpretation, assessment of relevant information, and selection of data to evaluate the rule are all features of the content-specific and content-general views, and the relevance theorists have done an admirable job of highlighting their importance. However, these processes are not where RT_{Wason} and the alternative content-specific accounts differ. What is at issue between these rival theories is (i) the content and form of the representations and inference procedures that compute equivalences and implications during the process of interpretation, (ii) the content of the representations thereby computed, and (iii) the content and form of the inference procedures that subsequently act on them. In RT_{Wason}'s account of content effects, the underlying representations are content-free logical formulae, the inferences embody the rules of logic, and concepts such as 'violate' or 'verify' are defined in content-free logical terms. In contrast, social contract theory and hazard management (precaution) theory posit content-specific representations and content-specialized proprietary inference procedures, some of which generate inferences not licensed by the rules of logic. These govern both interpretation and post-interpretive inferences. Moreover, these theories propose that when subjects are instructed to look for 'violations' in these domains, this activates content-specific concepts such as 'cheating' (illicitly taking a benefit) and 'endangerment', and *not* the content-free concept of violation found in the predicate calculus.

According to SCG (p. 53), "Subjects trust their intuitions, that is, the output of their spontaneous inferential abilities; without any further examination, they take the directly testable consequences that they have inferred in comprehending the rule to be the consequences through which the rule should be tested." In other words, relevance theory relies on spontaneous inference abilities to explain performance. If performance cannot be explained by invoking logical SIAs alone, then other SIAs

– including content-specific ones – need to be invoked. If content-specific SIAs, which supply non-logical inferences, need to be invoked to explain performance, then results on the selection task do indeed illuminate their presence and design. In the experiments below, we present several such cases.

2. Experiment 1: are logical connectives necessary?

Relevance theory postulates that performance on the selection task is guided by representations of logical form. Moreover, it is presented as a content-independent theory of discourse interpretation in which the logical forms computed reflect the structure of discourse *and not the structure supplied by domain-specific inferences and post-interpretive processes such as cheater detection*.

To see this, consider the narratives used by SCG to test RT_{Wason}. In them, the discourse is structured by the logic of denial. For example, the virgin mothers narrative has the following abstract form:

(a) Agent 1 (a credible source)[14] makes a surprising accusation: 'There are entities that are both *P* and *not-Q*!'
(b) Agent 2 (a questionable source) denies that this is true (he 'makes a joke out of these allegations'), and asserts instead that 'If an entity is *P* then it is also *Q*'.
(c) The narrator implies that Agent 2 is lying. You are asked which cards one needs to turn over to discover whether what Agent 2 said is true or false.

Note several things about this narrative structure. (i) In it, propositions are asserted, and then denied: basic functions of first order logic, which is content-independent. As a result, one can follow the logical structure of the narrative without knowing anything about what *P* and *Q* stand for. (ii) Given that Agent 2 denied the assertion made by Agent 1, discovering that Agent 2's claim is false is equivalent to discovering that Agent 1's claim is true. Consequently, all one needs to do is choose those cards that would confirm the assertion made by Agent 1 that 'There are entities that are both *P* and *not-Q*' (i.e. *P* and *not-Q*).[15]

The core of SCG's sweeping claim – that relevance theory 'explains the selection task' – depends upon the claim that performance is, in all cases and all domains, guided by representations of content-independent logical form unaided by other spontaneous inference abilities, such as social contract algorithms. To the extent that the logical structure of the discourse is altered by the removal of logical

[14] The identity of the accuser (Agent 1) is not directly specified, but the implication is that the accusations were made by journalists or other members of the community at large.

[15] Indeed, given this narrative structure, the conditional may have been superfluous. Making a joke out of allegations is itself a way of denying them. Given this, one wonders whether the same result might not have been obtained had Agent 2 simply said, 'I deny it!' – without further elaboration – in response to Agent l's assertion that there are entities that are both *P* and *not-Q*. If performance turned out to be the same when the conditional rule is absent, then it would be difficult to maintain that the virgin mother problem contributes much to our understanding of how people interpret or reason about conditional rules per se, or of when they deduce consequence (c). In that case, it would not be a good example of relevance explaining the selection task.

connectives, RT_{Wason} would predict that performance on the selection task should change: after all, with no logical connectives, there is no mechanism whereby the deductive consequences outlined in (a–c) in Table 1 could be triggered. So, a natural first test of whether relevance theory provides the explanation for social contract performance is to test RT_{Wason}'s prediction by replacing a conditional sentence, *If P then Q*, with two logically unconnected sentences, *P* and *Q*, and seeing what happens to subject performance.

Social contract theory, in contrast, has no commitment to logical formulae. It posits a domain-specific set of inferences that are automatically triggered when one represents a situation as involving social exchange (see Table 2) and a domain-specific representational system. This representational system is sensitive to the perceived costs and benefits of agents. Whether in narratives or in the language of thought, social interactions are typically described by invoking the intentions, desires, and beliefs of agents (Baron-Cohen, 1995; Dennett, 1987). When these are mapped onto the representational structure of a social contract, the interaction is recognized as involving social exchange and a cascade of domain-appropriate inferences are triggered. These inferences supply what SCG (pp. 87–88) refer to as the 'implied content' of conditionals in this domain. Table 3 shows the representational structure of a well-formed, sincere social contract in which goods are exchanged; baseline fraud (a form of deception distinct from cheating), hyperbole, irony, and other tropes can be understood as systematic deviations from the cost-benefit structure of a sincere social contract (Cosmides, 1985; Cosmides & Tooby, 1989).

Language is a useful means of acquiring knowledge of people's values in order to propose an acceptable deal, but a growing body of research on theory of mind suggests that humans possess many specialized psychological mechanisms for modeling the mental states of others, including what they might want, even on the basis of subtle cues like eye gaze (Baron-Cohen, 1995). Hence, social contract theory predicts that people should be able to engage in social exchange on the basis of such minimal cues, provided that the mutual representation of desired benefits is possible. If so, then even the fragmentary use of language should suffice for the purposes of engaging in an exchange. For example, lacking detailed knowledge of the native language, an anthropologist was able to propose a mutually comprehensible trade to Amazonian hunter-horticulturalists with as ungrammatical an utterance as 'One knife, two chickens' (Sugiyama, 1996). Experiment 1 attempts to recreate such a transaction when there is little shared language, so that logical connectives and conditional forms are omitted. It compares performance on a selection task in which there is an explicit rule – Villager to Farmer: 'If you give me some potatoes, then I will give you some corn' (*Conditional* version) – to a selection task in which no rule is stated, but individuals express what they want – Villager: 'I want some potatoes'; Farmer: 'I want some corn' (*Want* version). In these problems, the subject, who is cued into the role of the farmer (Gigerenzer & Hug, 1992), is asked to see whether any of the villagers have cheated the farmer (see Fig. 2).

2.1. Method

2.1.1. Subjects

For all three experiments reported herein, subjects were recruited in the student union building at the University of California, Santa Barbara. They were students and staff at the university. Upon completion of the experiment, all subjects were given a coupon redeemable for a snack on campus.

Sixty subjects were recruited for Experiment 1. They were randomly assigned to one of two groups: 30 to the *Conditional* condition and 30 to the *Want* condition.

2.1.2. Procedure and materials

All three experiments reported herein were conducted in a quiet room in the student union building. Subjects were recruited in the hallway outside the room and upon entering the room were seated, given a problem booklet, and asked to read the instructions and the problem carefully.

For all experiments, the problem booklet consisted of two pages. The first was a

You are a South American farmer. At the end of the harvest you find you have more potatoes than you need so you pack up some of them and travel to the neighboring village. When you get to the village four different people approach you, and though you don't speak the same dialect, you recognize that each of them is telling you,

Conditional Version	**Want Version**
"If you give me some potatoes, then I will give you some corn."	**"I want some potatoes."** You, in turn, know a little bit of their dialect, and tell them **"I want some corn."**

The cards below represent the four people who approached you. One side of the card tells whether or not you gave the person any potatoes and the other side of the card tells whether or not that person gave you any corn.

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

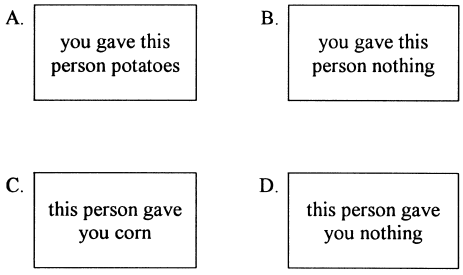

Fig. 2. Selection tasks used in Experiment 1. The only differences between the *Conditional* and *Want* versions of the task are indicated in the boxes, which are placed where the alterations occurred in the task scenarios (e.g. the *Conditional* version read '…you recognize that each of them is telling you, 'If you give me some potatoes, then I will give you some corn.' The cards below…'). The boxes did not appear in the versions of the task given to subjects.

standard Wason task instruction page (Cosmides, 1985, 1989). The second page contained a Wason selection task.

In Experiment 1, the Wason selection task on the second page described an unfamiliar scenario in which a South American potato farmer is taking his excess produce to the neighboring village (see Fig. 2).

Subjects received one of two versions of the selection task. In the *Conditional* version the farmer was offered a deal in a conditional statement *'If you give me some potatoes, then I will give you some corn'*, whereas in the *Want* version the people in the neighboring village tell the farmer *'I want some potatoes'* and he in turn tells them *'I want some corn'*. The two versions of the problem are otherwise identical.

2.2. Predictions

2.2.1. Social contract theory

Without the use of any logical connectives in their speech, the participants in a social exchange ought to be able to propose a deal simply by expressing what they want from the other party. Regardless of the means employed in reaching agreement, once a deal is agreed to, cheating is possible. Social contract theory therefore predicts that the dominant response will be to choose the *benefit accepted* card ('You gave this person potatoes') and the *requirement not satisfied* card ('This person gave you nothing') on both versions: when the deal is phrased as a conditional *and* when it is struck via two atomistic phrases, each expressing what one of the parties wants. Although the *Conditional* version contains more cues specifying that the situation involves exchange than does the *Want* version (e.g. the *Conditional* version talks about 'giving', and includes an explicit statement that an act of giving is contingent upon having received a good, specifies who is giving what, and so on), performance levels should nevertheless be similar in the two versions.

2.2.2. Relevance theory

2.2.2.1. Want version Relevance theory posits that logical equivalences triggered by representations of content-independent logical form are sufficient to explain performance on the selection task. It is therefore instructive to consider the *Want* narrative in its content-free form. Abstracting from the specific content, one can see that the logical form of the discourse is as follows: *Each villager says 'P' to the farmer. The farmer replies 'Q'. See if any of the villagers have cheated the farmer*. This version has no logically correct answer: no rule of the form '*If P then Q*' has been stated, so one cannot generate logical equivalents of this rule, or look for logical violations of the rule or its equivalents. For this reason, RT_{Wason} does not predict a high level of '*P & not-Q*' responses on the *Want* version. Performance should be random; alternatively, one could argue that subjects' attention will be drawn to the *P* card and the *Q* card, since these are the propositions mentioned in the narrative.

To maintain that relevance theory 'explains' the selection task is to maintain that one can explain performance *without* having to invoke social contract algorithms or

other contentful spontaneous inference abilities. It will not do, then, to say that two agents stating what they want causes subjects to infer that a deontic social contract rule is in place, and that subjects then reason from this implicit rule. On such an account, social contract algorithms are causing the crucial inference to be made and are, therefore, doing the explanatory work. Nevertheless, even if one were to accept such a proposal to make an RT_{Wason} account of this experiment possible, RT's commitment to logical inferences and/or equivalences means that as many people would answer '*not-P & Q*' as '*P & not-Q*', as will be explained later (see Section 2.3).

2.2.2.2. Conditional version SCG argue that relevance theory predicts high levels of the logically correct '*P & not-Q*' response on selection tasks employing deontic rules, because they are interpreted as forbidding cases of *P-and-(not-Q)* (see below for explanation). Although the rule we tested (*Conditional* version) does not use deontic operators such as 'must', 'may', 'obligated' or 'entitled', a context in which one is asked to look for cheaters suggests that the rule be interpreted deontically as 'If you give me some potatoes, then I will [be obligated to] give you some corn'. Hence, faced with an explicit deontic rule, SCG would expect most subjects to choose *P* ('You gave this person potatoes') and *not-Q* ('This person gave you nothing').

Although SCG explain high performance on deontic rules by maintaining that their communicative intent is to forbid cases of *P-and-(not Q)*, we would like to point out that the reasoning on which this prediction is based does not extend easily to the class of private exchanges (trades) as distinct from social laws. In formulating their argument that the communicative intent is to forbid, SCG analyzed social laws, where their argument more plausibly fits, such as the drinking age rule: 'If a person drinks beer, then that person must be over 19 years old' (Griggs & Cox, 1982). But social contracts expressing a trade lack key features upon which SCG's analysis of deontic rules is based. Their purpose is to enable mutual access to goods through a trade, *not* to forbid access to a good, and this has implications for what subjects should find relevant. As we will explain below, a consistent application of relevance theory to private exchanges should lead subjects to answer '*P & Q*' or '*P, Q, & not-Q*' for trades, rather than '*P & not-Q*' (see Section 2.3 for explanation).

2.3. Results and discussion

As predicted by social contract theory, most subjects chose the cards relevant to detecting cheaters in both versions. In the *Conditional* version, 67% of subjects selected the *benefit accepted* card (*P*) and the *requirement not satisfied* card (*not-Q*) and no others. Likewise, 50% of the subjects in the *Want* version picked all and only these two cards, even though this version lacked any logical connectives. The difference between these two conditions was not significant (test of proportions, $Z = 1.31$, $P > 0.09$). In other words, a change that is dramatic on any theory emphasizing logical form – the removal of all logical connectives – made no significant difference to subjects: about as many gave the fully correct cheater detection response when logical connectives were absent as when they were present.

There was a slight (but non-significant) tendency for subjects to choose all and only the two cheater detection cards more often in the *Conditional* than in the *Want* version. But this is to be expected: subjects, who must infer the type of interaction (social exchange, threat, warning, self-expression, and so on) on the basis of cues, should do so more reliably when there are more cues, as in the *Conditional* version than when there are few cues, as in the *Want* version. But this difference isn't as informative as the absolute level of *benefit accepted* and *requirement not satisfied* card selections. Fifty percent of the subjects in the *Want* condition picked these cards – a level of performance that is far above chance (6.25%) and similar to performance in the *Conditional* version.

The full pattern of card selections is given in Table 4. For the reader's convenience, in the *Want* version we have called the 'You gave this person potatoes' card the *P* card and the 'This person gave you nothing' card the *not-Q* card.[16] The second most common response on both versions was to choose the *P* card alone (*Conditional*: 20%; *Want*: 13.3%). Moreover, there was a significant positive correlation between the rank order of selections in the two conditions ($r_s(8) = 0.68$, $P < 0.025$, one-tailed), suggesting no difference in the overall pattern of selections in the two conditions.

These results are not consistent with the predictions of RT_{Wason} – that is, in its form unsupplemented by social contract theory. This is most striking for the *Want* version, in which no conditional rule was stated. If one excludes interpretations of RT_{Wason} that depend on the existence of social contract algorithms to do the interesting inferential work, then answers should either be randomly distributed – which implies a '*P & not-Q*' response rate of 6.25%, not 50% – or the predominant response should be '*P & Q*', as these are the values mentioned in the narrative. As one can see from Table 4, however, only 6.7% of subjects answered '*P & Q*', which does not differ from chance.

What if one interprets relevance theory more liberally? Suppose one admits that social contract algorithms cause the interaction of the villagers and the farmer to be interpreted as an exchange – as an implicit conditional – but then claims that a content-independent discourse comprehension module takes over from there? Have the 50% of subjects who chose the 'You gave this person potatoes' card and the 'This person gave you nothing' card made the logically correct choice?

That conclusion would be unwarranted. RT_{Wason} claims that performance on the selection task can be explained by a discourse comprehension module that applies content-independent *logical* equivalences, thereby generating *logically* correct card choices in response to deontic conditional rules. A device of this kind would have to interpret the narrative – including the implied conditional – using logical concepts rather than social contract algorithms. It would have to map the word 'cheating' onto a more general notion of 'violation'. So let's say, as SCG (p. 85) themselves suggest, that this device interprets the instruction 'to see if any of these people have cheated you' as an instruction 'to see if any of these people have violated a rule', where

[16] Of course there are no logical connectives in the *Want* version that define particular propositions as either antecedents or as consequents. But the P, Q notation is simpler for the reader to follow than a table full of potatoes, corn, and nothing.

Table 4
Experiment 1: pattern of card selections (percentage of subjects)

Pattern	Condition	
	Conditional	*Want*
P, not-Q	66.7	50.0
P	20.0	13.3
P, Q	3.3	6.7
not-Q	3.3	6.7
P, not-P	0.0	6.7
None	3.3	6.7
Q	0.0	3.3
P, Q, not-Q	0.0	3.3
not-P, not-Q	0.0	3.3
All	3.3	0.0

'violation' is defined logically as the co-occurrence of a true antecedent and a false consequent. What cards would the subject then choose?

The cards chosen necessarily depend on *which* out of several possible implicit conditionals each subject inferred. Remember, this liberal account requires a specific set of spontaneous inference abilities – those provided by social contract algorithms – to infer a conditional. But the grammar of social exchange (Cosmides, 1985; Cosmides & Tooby, 1989) allows the proposed trade to be expressed in several different ways. Indeed, given the fragmentary information in the *Want* version, there are four equivalent ways of expressing an exchange rule:

(1) Villager: 'If you give me some potatoes, then I will give you some corn.'
(2) Farmer: 'If you give me some corn, then I will give you some potatoes.'
(3) Farmer: 'If I give you some potatoes, then you must give me some corn.'
(4) Villager: 'If I give you some corn, then you must give me some potatoes.'

To avoid confusion, let us define the proposition 'You gave this person potatoes' as *P* and the proposition 'This person gave you corn' as *Q*. The logically correct answer is always to choose the cards representing a true antecedent and a false consequent. For the conditionals expressed in (1) and (3), the true antecedent is the 'You gave this person potatoes' card – *P* – and the false consequent is the 'This person gave you nothing' card – *not-Q*. These are the cards most subjects chose. But for (2) and (4), the true antecedent is the 'This person gave you corn' card – *Q* – and the false consequent is the 'You gave this person nothing' card – *not-P*. Thus, '*P & not-Q*' is the logically correct answer to (1) and (3), but '*Q & not-P*' is the logically correct answer to (2) and (4).

All four ways of expressing the implied exchange are pragmatically felicitous.[17] Moreover, there is nothing in the narrative that would make the construction of one

[17] Although all four are pragmatically reasonable, some might find (1) and (2) slightly more felicitous than (3) and (4). This makes no difference to the argument though. Even if we restricted the set to (1) and (2), the logically correct answer to (2) is '*Q & not-P*'.

of these conditionals more probable than any of the others. The best we can do is assume that they are equally likely. Under these circumstances, the best RT_{Wason} prediction is as follows: assuming that the *Want* version will elicit about as many logically correct responses as the *Conditional* version did – 67% – half of these subjects (about 33%) will answer '*P & not-Q*' and half will answer '*Q & not-P*'. This did not happen. Whereas 50% of subjects in the *Want* version answered '*P & not-Q*', none chose '*Q & not-P*'. Thus, performance in the *Want* version lends little support to this watered down version of relevance theory. Even if RT_{Wason} allows social contract algorithms to do the initial, interpretive work, *a post-interpretive cheater detection procedure must be invoked to explain the results*. This is because subjects chose the cards relevant to cheater detection, no matter which logical categories they fell into. Cheater detection procedures short-circuited the application of logical SIAs and the search for logical violations.

An even more liberal interpretation of RT_{Wason} would be that social contract algorithms cause one to interpret the interaction as involving exchange, allow you to understand what counts as cheating, and define cheating as illicitly taking a benefit (rather than the more general notion of violation from logic). But on this interpretation, social contract algorithms are doing all the explanatory work: it is not clear what introducing relevance theory adds to the explanation.

Indeed, as we discuss below, SCG's explanation for the results of the perspective shift experiments by Gigerenzer and Hug (1992) suffers from exactly the same difficulties.

2.3.1. Relevance theory does not explain perspective shifts

According to social contract theory (Cosmides & Tooby, 1989), what counts as cheating is perspective-dependent (e.g. Farmer cheats by accepting corn and not giving potatoes; Villager cheats by taking potatoes without giving corn), and social contract algorithms should make one good at detecting cheating from the perspective of either party to an exchange. Gigerenzer and Hug (1992) tested these predictions by giving subjects rules such as 'If an employee works on the week-end, then he gets a day off during the week'. They found that subjects cued into the role of an employee chose the 'worked on week-end' and 'no day off during week' cards (*P* and *not-Q*; logically correct), which are the cards relevant to seeing whether an employer had cheated an employee. In contrast, subjects cued into the role of an employer chose the 'did not work on week-end' and 'took day off during week' cards (*not-P* and *Q;* logically incorrect), which are the cards relevant to seeing whether an employee had cheated an employer. This is a straightforward prediction of the theory that one has cheater detection procedures and social contract algorithms that derive the (various) implications of an agreement to enter into a social contract from the surface content of utterances.

In (apparent) contrast to this explanation, SCG (pp. 86–88) claim that subjects cued into the employer role answer '*not-P & Q*' because:

(i) they *interpret the rule as a reciprocal contract*, then
(ii) they infer the 'implicit content' of the rule *that is relevant to judging what*

counts as cheating from the employer's perspective (this being *If Q then P*), then (iii) logical equivalences are derived from this *inferred* rule to settle on what would count as a (logical) violation (i.e. *Q-and-(not-P)*).

But what procedures allow the 'week-end' rule – and not other conditional rules – to be interpreted as a 'reciprocal contract'? Certainly not logical procedures. And what inference procedures allow the appropriate implicit content to be inferred? The only conditional that *logically* follows from the stated one is 'If an employee does not get a day off during the week, then he does not work on the week-end' (because *If P then Q* implies *If not-Q then not-P*). Neither the original statement nor this one says anything about what an employer would view as cheating. To derive implicit content pertinent to *this* issue, one needs the (non-logical) inferences supplied by social contract algorithms *and* their definition of cheating (to tell which of the many implications allowed by the social contract algorithms is relevant to the issue of an employer being cheated (see Table 2, especially item 8;[18] also Table 3)). This is the only explicit procedure we know of that would allow one to (i) derive 'If an employee takes a day off during the week, then he must work on the week-end' (i.e. *If Q then P*) as (an) implication of the rule given, and (ii) realize that this is the implication that is most pertinent to the issue of employers being cheated, thereby (iii) allowing the further derivation of *Forbid*(*day off during week* and *no work on week-end*) as a logical equivalence.

This explanation is neither parsimonious nor content-general. The deployment of logic – which is content-independent – is what lends RT_{Wason} its generality.[19] Yet SCG's explanation does not rely on logic alone. The subject must first make an invalid inference – that *If P then Q* implies *If Q then P*[20] – and then derive a logical equivalent (*Forbid(Q-and-(not-P))* of this *invalidly* inferred statement (and *not* from the original rule!). Second, because there are many possible invalid inferences, SCG's account requires a content-independent procedure whereby the subject succeeds in settling on the *appropriate* invalid inference. But logic will not solve this problem either. One needs to invoke social contract algorithms to solve these

[18] The rule was given in the third person, allowing subjects to switch their roles. When the employer is the 'I' of Table 2, the rule may be translated to 'If you give me a day of work on the week-end, then I will give you a day off during the week'; by #8 this entails 'If you accept a day off during the week from me, then you are obligated to give me a day of work on the week-end'.

[19] One could claim that the notions of cognitive effect and cognitive effort also lend it generality. However, cognitive effects are, to a substantial extent, a function of inferential power, which comes from spontaneous inference abilities, many of which are domain-specific. To the extent that differential performance is a function of different cognitive effects – holding all else constant – then the selection task can indeed illuminate domain-specific inference abilities. Moreover, cognitive effort is also heavily shaped by the presence or absence, in the mind, of evolved specializations designed to automate various inferences. Also, as we will argue below, some seemingly general principles of cognitive effort (such as processing negations) may actually be domain-specific.

[20] We should point out that social contracts are not biconditionals. A person who has met the requirement but not taken the benefit has not cheated. To be a biconditional, there would have to be two violations: (1) *benefit accepted* and *requirement not met*; and (2) *benefit not accepted* and *requirement met.*

problems: making the crucial inferences ((i) and (ii) above) requires their non-logical rules of transformation and their content-dependent definition of cheating.

So RT_{Wason}'s explanation for perspective shifts does not eliminate the need for social contract algorithms; indeed, it depends on them. Moreover, the tortuous procedure RT_{Wason} requires – introducing logical equivalences after a dense series of non-logical implications have been derived by social contract algorithms during the interpretive process – is what artificial intelligence researchers would call a kludge. After all, the same cheater detection procedure needed to select *If Q then P* as the correct interpretation is sufficient for detecting the appropriate violation. On social exchanges, RT_{Wason}'s logical SIAs are superfluous.

Social contract theory provides a simpler and more elegant explanation. Table 3 gives the social contract algorithms' rules for translating an exchange into the value system of each party. Both translations are automatically computed (this is what it means to say that people *understand* the utilities involved in an exchange). 'Work on week-end' is assigned the value $C_{employee}$; 'day off during week' is assigned the value $B_{employee}$. Cueing the employer's perspective activates procedures for detecting employee cheating, which are applied directly to the employee-values representation of the rule (i.e. If $C_{employee}$ then $B_{employee}$). This procedure 'looks for cheaters', that is, it checks any situation in which an employee has taken the benefit ($B_{employee}$ card) and any situation in which an employee has not paid the cost (not-$C_{employee}$ card).

2.3.2. Relevance theory makes different predictions for social laws and for trades

More significantly, an analysis of how RT_{Wason} applies to personal trades calls into question whether performance on even the *Conditional* version of Experiment 1 supports relevance theory. SCG did not test RT_{Wason} on deontic rules, but they claim that, with minor modifications, their explanation of performance on descriptive rules (ones not involving denial) can be extended to explain performance on deontic rules. But their argument, which is formulated using the drinking age and cholera rules, runs into problems when applied to a rule expressing a trade. This point is important, because it suggests that all deontic rules are not the same, and that RT_{Wason} cannot provide a general explanation for results with deontic rules.

Assume a pragmatic context in which no prior assertions have been made about what features co-occur with *P*. SCG argue (quite reasonably) that in this context, when a speaker utters a descriptive rule of the form '*If P then Q*', it is to inform hearers – contrary to their expectation – that (i) something with the feature *P* will also have the feature *Q* (inference (a)), and (ii) cases of *P-and-Q* do exist (inference (b); see Table 1, Panel A). There would be no point in asserting this unless hearers had previously assumed that cases of *P-and-(not-Q)* are either possible or actually exist, so inference (c) lacks relevance: it does not lead you to infer anything about the world that you did not already believe to be true. (Making inference (c) is also effortful, because it involves processing double negatives ('contradict', '*not-Q*').) As a result, subjects correctly choose the *P* card (inferences (a) and (b)), incorrectly choose the *Q* card (inference (b)), and fail to choose the *not-Q* card (inference (c)).

SCG present a similar analysis of the drinking age problem (Griggs & Cox, 1982), a Wason selection task employing the following rule: 'If a person drinks beer, then that person must be over 19 years old'. The parallels are drawn in Table 1, Panel B. According to SCG (pp. 84–86), the drinking age problem differs from descriptive conditionals in three key ways. As a result, it elicits analogs of inferences (a) and (c), but not of (b), causing the majority of subjects to choose *P* and *not-Q*, rather than *P* and *Q*. The three differences are as follows:

(i) *Inference (b″) is not made*. Most people's default assumption – their assumption about what would be happening if the drinking age rule did not exist – is that there will be some instances of adults drinking beer. Thus, the rule cannot have been made to inform you that cases of *P-and-Q* (adult beer-drinkers) exist. Similarly, knowing that adult beer-drinkers already exist, it is unlikely that the purpose of the rule is to create instances of adult beer-drinkers by *requiring* their existence. Thus, people are unlikely to make inference (b″) in Table 1 (Panel B), and, therefore, will not be tempted to choose the *Q* card along with the *P* card. (In this way, it is unlike a descriptive rule, which is stated to inform others that, contrary to their prior belief, cases of *P-and-Q* do exist – i.e. inference (b).) Moreover, finding adults drinking beer "is a trivial event from which nothing significant follows", so looking for instances of *P-and-Q* would produce no cognitive effects. This further reduces the likelihood of (b″) being drawn and *Q* being chosen.

(ii) *Inferences (a″) and (c″) are made*. Since the rule could not have been stated to inform one that instances of *P-and-Q* exist, an assumption of relevance implies that there must be another reason that the rule was made, so one looks for a logical equivalence that is relevant given one's background assumptions. This leads to inference (c″): the rule was made to *forbid* certain individuals (underage ones) from drinking beer, presumably to preserve safety or public morality. In other words, the purpose of the drinking age rule is to forbid the co-occurrence of *P-and-(not-Q)*, thereby ensuring that if *P* cases (beer-drinkers) exist they will also have the feature *Q* (adulthood) (i.e. inference (a″)). SCG see the *Forbid[P-and-(not-Q)]* interpretation of deontic conditional rules as analogous to the *Deny[P-and-(not-Q)]* interpretation of descriptive conditional rules involving denial.[21] Moreover, finding underage beer-drinkers has more cognitive effects than finding adult beer-drinkers: even in the absence of the rule, young people drinking alcohol is "seen as a social or moral problem and is therefore more significant". Given that the rule does exist, its violation could lead to sanctions. Therefore, finding instances of *P-and-(not-Q)* carries more implications (supports more inferences).

(iii) *Violations are easy to represent*. On the cognitive effort side, because there are no 'nots' involved in representing *not-Q*, it is just as easy to represent cases of *P-and-(not-Q)* (16-year-old beer-drinkers) as cases of *P-and-Q* (25-year-old beer-drinkers). As a result, (c″) is easier to infer than (c). Indeed, it is argued that the *P-*

[21] It is not clear why. Logic has to do with truth, not with prohibition. To the extent that SCG wish to define the selection task as involving the search for data relevant to determining whether statements are true or false, then this theoretical move is unwarranted.

and-(not-Q) case may be even easier to represent than the *P-and-Q* case, insofar as it is linguistically highlighted by the lexicalization of concepts such as cheating and violation. SCG claim that, in the deontic domain, "there is a wealth of terms to designate violators and violations of specific rules or types of rules, for example, adultery, arson, blasphemy…[but]…no correspondingly rich terminology to designate non-adultery, non-arson, non-blasphemy, etc. …there are no counterpart one-word labels for the many varieties of rule-abiding, or even for rule-abiding in general. *P-and-(not-Q)* events are instances of at least one lexicalized concept available to subjects: that of violation, whereas, in general, *P-and-Q* events don't belong to a distinctive named category."

The first two of these assumptions/arguments may hold for certain kinds of social laws (we think the third is problematic even for social laws; see below). But *none* of the three hold for a private exchange (i.e. an agreement to swap or trade). Unlike (some) social laws, trades are not proposed to *Forbid[P-and-(not-Q)]*, i.e. cheating. Instead, the communicative intent of a person proposing a trade is to create an exchange – to create a situation of mutual benefit. This creates quite different assumptions of relevance, which should trigger different spontaneous inferences and card choices if RT_{Wason} is correct. Unlike social laws:

(i) *Trades support inference (b'').* Default assumptions for trades are the opposite of default assumptions for social laws. For social laws, people assume that, *in the absence of the law*, *P* will occur (some people will drink beer) and there will be instances of *P-and-Q* (adult beer-drinkers). Because instances of *P-and-Q* already exist, creating such instances cannot be the law's intended purpose. But for trades, people assume that, *in the absence of an agreement to trade*, *P* will *not* occur, and there will be *no* instances of *P-and-Q*. Consider the communicative intent of the villagers who propose a trade. The villagers propose the conditional rule – they offer corn in exchange for potatoes – precisely because they believe that nothing will happen unless they do. They will get no potatoes (*not-P* will occur), they will give no corn (*not-Q* will occur) and, therefore, there will be no instances of *P-and-Q* – of them getting potatoes and the farmer getting corn.

The villagers' purpose in offering an exchange is exactly *to create* instances of *P-and-Q*. In proposing the conditional rule, that was their communicative intent. If both parties agree to the rule, "the rule, together with a presumption of relevance, requires in most contexts that there be cases of P and Q" – i.e. inference (b''). In the farmer problem, all parties did agree to the rule, so subjects should expect some trades to take place. Moreover, finding instances in which people cooperate – in which they trade fairly – is an important event, from which a great deal of significance follows. Honest traders are people from whom you can benefit, and with whom you should want to forge relationships – perhaps lifelong trading relationships. When, on entering a new situation, you discover an instance of *P-and-Q*, you have found *two* individuals who kept their word and behaved in an honorable manner, so looking for such instances should produce large cognitive effects. This increases the relevance of *P-and-Q*, and the likelihood of drawing inference (b''). This is the inference that leads to the choice of the *Q* card (and the *P* card).

(ii) *Trades do not support inference (a″), and only weakly support (c″)*. For the social law, because people assume that cases of *P* will exist no matter what (i.e. there will be beer-drinkers, even without the law), it is reasonable to assume the rule was created to prevent wanton beer drinking: that is, to ensure that beer-drinkers (*P*) are always adults (*Q*) (inference (a″)). This is unlike a trade. In saying, 'If you give me some potatoes, then I will give you some corn', the villager is *not* trying to prevent the wanton distribution of potatoes by an altruistic farmer. The villagers assume that the farmer will *not* be giving them potatoes in the absence of an exchange agreement, i.e. their default assumption is that cases of *P* will *not* exist. So the purpose of the rule cannot be that implied by inference (a″): to ensure that, if *P* cases happen to exist (i.e. farmer starts giving away potatoes), they will also have the feature *Q* (farmer will get corn). Presumably, the villagers would be delighted to receive free potatoes (and also surprised if the farmer started handing them out for free). They have no interest in *requiring* that a willing benefactor be rewarded with corn. Rather, their goal is to *induce* the farmer to give potatoes by *promising* to reciprocate with corn: to change the world from *(not-P)-and-(not-Q)* to *P-and-Q* (to capture the difference, one would have to create a modified version of (a″) that captures the complex intercontingencies listed in Table 2).

Furthermore, given these default assumptions, considerations of relevance do little to promote inference (c″) (which leads to the selection of the *not-Q* card). *P-and-(not-Q)* represents an instance in which a *villager* has cheated: he gets potatoes but gives no corn. People offer trades in order to get things, not to forbid *themselves* from cheating. Therefore, in uttering the conditional rule proposing a trade, the villager's intent cannot be to *Forbid[P-and-(not-Q)]* – to prevent himself from cheating. In other words, the deductive consequence in (c″) cannot be the intended meaning of the conditional. This is quite different from the case of the social law.

Of course neither party wishes to be cheated, but this is a secondary consideration, not the speaker's communicative intent. To avoid a combinatorial explosion of implications, RT_{Wason} has a stopping rule: to determine the speaker's intent, follow a path of least effort and settle on the first implication that satisfies your expectations of relevance. Clearly, the villager's intent in proposing the rule and the farmer's in accepting it was to create an exchange – not to forbid cheating. The villager is *promising Q* if *P* happens, to induce the farmer to give him potatoes. At best, one might argue that promising *Q* (if *P*) secondarily entails a promise to not cheat.[22] But this is a more convoluted inference than (c″): it is *Promise[not-[P-and-(not-Q)]]* which should be effortful to represent (because of the embedded nots) and is not one of SCG's three deductive consequences. Even so, it is a

[22] Indeed, it might, as a secondary matter, entail many things, e.g. a promise to not murder the farmer before the transaction, to not steal his potatoes, to not damage the corn before delivery, and so on. But are we to assume that *all* of these consequences are spontaneously deduced? If not these, then why the promise to not cheat? After all cheating is only one way – out of many – in which one can fail to deliver on the promise of corn for potatoes.

downstream entailment, past the stopping rule: the promise of *Q* is the inducement, not the promise to not cheat.

Does finding instances of *P-and-(not-Q)* have more cognitive effects than finding instances of *P-and-Q*? Certainly it is relevant to discover that an individual has cheated: this will affect your decision to deal with that person in the future. But so is the discovery that an individual was honest. Indeed, discovering that a villager cheated arguably has just one implication: you should not trade with this person in the future. But discovering that a villager was honest has many and continuing implications for the possibility of a long and mutually beneficial relationship. So the production of cognitive effects does little to promote the inference of (c″) compared to (b″). The arguments of Cosmides and Tooby (1989, 1992) for why a cheater detection mechanism should have evolved do not depend on the production of cognitive effects. Indeed, to make RT_{Wason} testable, one can (and should) go beyond speculation on cognitive effects. In a series of experiments, it appears as if subjects find honest behavior more relevant in making subsequent inferences about potential cheating than exposure to instances of cheating (Cosmides, Tooby, Montaldi & Thrall, 1999).

(iii) *Trades are easier to represent than violations.* What about cognitive effort? For the trade problem, 'This person gave you nothing' (*not-Q*) involves a not-so-implicit 'not' (i.e. 'no-thing'). This should be more difficult to represent than 'This person gave you corn' (*Q*). Thus, the *P-and-(not-Q)* case (cheating) should be more difficult to represent than the *P-and-Q* case (trading). As a result, (c″) should be more effortful to infer than in the social law (the drinking age problem).

SCG argue that lexicalization of a concept helps highlight its importance. But both concepts – cheating and exchanging (trading, swapping, etc.) – are lexicalized in English. Indeed, the lexicalization issue cuts both ways.[23] For exchanges, there is a rich vocabulary for representing the *P-and-Q* case: analogous to saying someone cheats is to say they are honest, trustworthy, honorable, fair, cooperative, or reliable. The fact that these words do not appear in the story is not to the point: SCG do not require that the word 'cheater' be present either, as they believe the concept will be inferred from less specific terms, such as 'violate' (SCG, pp. 42–43). If the context is sufficient to trigger the idea of cheating, then surely the fact that a deal is being offered and accepted is sufficient to trigger the simple idea of an honest exchange – of trading. So the lexicalization argument is not sufficient to highlight the *P-and-(not-Q)* case over the *P-and-Q* case.

To what extent is inference (c″) helped by the fact that subjects are specifically asked to look for cheating? If RT_{Wason} is to provide an account that is "wholly general", then the answer has to be 'not much'. Violation – like cheating – is lexicalized in English, and there are many selection tasks using ordinary descrip-

[23] The same holds for social laws. Contrary to SCG's claim, cases in which a person is rule-abiding are in fact lexicalized in English. It is as natural to call a person 'law-abiding' as 'law-breaking'; it is as natural to say 'he obeyed/followed the rule' as 'he broke/violated the rule'. And a rich vocabulary for social rule following does in fact exist: blasphemy/piety; adultery/fidelity, chastity; treason/loyalty; to cheat/to honor or to cooperate, etc.

tive rules in which the subject is specifically asked which cards one would need to turn over *to see if any of them violate the rule* (rather than asking which cards one needs to turn over to determine whether *the rule* is true or false – a more complex issue). But specifically asking subjects to look for violations is not sufficient to trigger inference (c) for these descriptive rules (see Experiment 2).

Taken together, what do (i–iii) mean for card choices on conditional rules expressing a trade? Subjects should not make inference (a″), they may or may not make inference (c″), and they will definitely make inference (b″). Contrary to the social law case, inference (b″) is highly relevant for a trade, which should focus attention on cases of *P-and-Q*. As a result, most subjects should choose the *P* card and the *Q* card (whether or not they also choose the *not-Q* card). This does not happen. Indeed, only one subject answered '*P & Q*' in Experiment 1's *Conditional* version, which expressed a trade (and none answered '*P*, *Q*, *& not-Q*'). Given that an RT_{Wason} analysis of trade leads to the inference that the purpose of proposing a trade is to create cases of *P-and-Q*, it is difficult to understand how subjects could omit the *Q* card. But over 93% of subjects did just that. Indeed, although the 67% of subjects who answered '*P & not-Q*' gave the logically correct answer, this is not an answer that makes sense given a careful, relevance theoretic analysis of the problem. Conversational pragmatics plus logic should cause subjects to choose the *Q* card. Therefore, failing to select the *Q* card results in an answer that is incomplete and contrary to what RT_{Wason} predicts, in exactly the same sense as failing to select *not-Q* would result in an incomplete answer to a social law problem. The fact that people do select the cheater detection cards, but do not select the only card that is relevant to a trade without also being relevant to cheater detection, is what one would expect if a cheater detection circuit were short-circuiting relevance effects.

Will people infer – through considerations of relevance alone – (c″) from a trade, which leads to choosing the *not-Q* card? Trades are like ordinary descriptive rules, in that their purpose is to create a link between *P* and *Q*, so RT_{Wason} predicts that the *P* card and the *Q* card choices should be common, and suggests that SCG should not predict any more *not-Q* choices for trades than one finds in the descriptive case. Still, cases of cheating – while not as fraught with implication as cases of honest exchange – might be thought to produce more cognitive effects than violations of arbitrary descriptive rules, perhaps increasing *not-Q* choices somewhat over the descriptive case. This would lead to more '*P, Q, & not-Q*' answers than one finds on ordinary descriptive rules. Yet no subjects in the *Conditional* trade condition gave this response.

Note, moreover, that a consistent application of RT_{Wason} leads to the prediction that the *not-Q* card *should be chosen less often for trades than for social laws*. The case is parallel to SCG's explanation of performance on ordinary descriptive rules versus ones involving denial, in which they argued that the same sentence, 'If P then Q', can be represented either as $\exists x\ (Px\ \&\ Qx)$ or as *Denied*-$[\exists x\ (Px\ \&\ not\text{-}Qx)]$, leading to more *not-Q* choices in the latter case. The deontic domain is similar. People agree to trades to create cases of *P-and-Q*, whereas people institute social

laws to forbid cases of *P-and-(not-Q)* (at least according to SCG),[24] so laws should elicit more *not-Q* selections. The wish of both parties to avoid being cheated is, at best, a downstream constraint rather than the purpose of making an agreement to exchange. (Note: to the extent that subjects were sensitive to this downstream constraint, but otherwise guided by principles of relevance, they should select cards corresponding to cheating by *both* parties – villager cheats ('*P & not-Q*') *and* farmer cheats ('*not-P & Q*') – because both produce cognitive effects. Yet *not-P* – like *Q* – was rarely chosen.)

Experiment 1 did not compare performance on a trade to performance on a social law, but Cosmides (1989) did, for closely matched contents (Experiment 1, law; Experiment 2, trade, where *requirement not satisfied* corresponded to the logical category *not-Q*; Experiment 3, law; Experiment 4, trade, where *requirement not satisfied* corresponded to the category *not-P*). Because these experiments were not designed to test relevance theory,[25] we do not want to overanalyze them here. We simply note three things. First, the tendency to choose '*P & Q*' was no higher in the trade problems than in the law problems (the *requirement satisfied* card (*Q* in Experiments 1 and 2, *P* in Experiments 3 and 4) was never chosen when it was *Q*, and almost never chosen when it was *P*). Second, there was no significant difference in cheater detection for laws versus trades (i.e. in choosing all and only the *benefit accepted* and *requirement not satisfied* cards). Moreover, there was no greater tendency to choose the *requirement not satisfied* card – crucial for cheater detection – regardless of other card choices in the law than the trade problems (i.e. law, Experiments 1 and 3; trade, Experiments 2 and 4).

2.4. Conclusions from Experiment 1

In our analysis of Experiment 1 and related results from the existing literature, we have tried to apply RT_{Wason} very carefully. We conclude from this analysis that RT_{Wason} can provide no explanation for trades or for perspective shift experiments (most of which also involve trades) that are free of social contract algorithms.

[24] It is important to note that SCG's framing is only one way of analyzing a social law involving benefits (such as drinking beer). The same analysis we have provided for trade we also applied to the drinking age and similar problems, allowing both sets of results to be parsimoniously explained with one theory. On social laws that elicit high performance, the potential cheater is seeing a situation that, from her point of view, has exactly the properties of any other social contract: a rationed benefit access to which has been made conditional on her meeting a requirement. Also, like the villager who proposes the rule because he wants to get potatoes, the people who proposed the drinking age law also wanted a good: safer streets. People restrict access to benefits to create a *situation* that benefits them. That situation can be that which is obtained when access to a good is restricted. This is why Cosmides and Tooby (Cosmides, 1985, 1989; Cosmides & Tooby, 1989; Tooby & Cosmides, 1996) have always considered social laws of the form, 'If you take the benefit, then you must satisfy the requirement' to be instances of social exchange. This is the only theory we know of that accounts for why not all social laws elicit high performance, but ones that have the form of social contracts (or the form of precautions) do (see also Appendix A).

[25] Because this was not the purpose of the experiments, we were not trying to match them for pragmatic felicity.

Indeed, social contract algorithms are required both for the interpretive process *and* for the post-interpretive inferences that guide card selection.

SCG's attempt to explain the Gigerenzer and Hug (1992) perspective shift experiments relied entirely on social contract algorithms during the interpretive process to make a complex series of inferences that cannot be made using logical equivalences alone. After the rule has been re-interpreted via social contract algorithms, SCG posit that logical equivalences, rather than a cheater detection subroutine, are applied. Yet the same cheater detection subroutine that is sufficient to make the correct card selection during this last step has to be invoked earlier for the interpretive process to work.

Even if one were to accept this inelegant solution to the perspective shift experiments, it is not sufficient to explain the high levels of '*P & not-Q*' responses in the *Want* condition of Experiment 1. Again, RT_{Wason} would have to heavily invoke social contract algorithms in the interpretive process, but then apply logical equivalences rather than a cheater detection algorithm to the interpreted rule. But this process would have produced as many '*not-P & Q*' choices as '*P & not-Q*' choices, and this did not happen. To explain why subjects chose the correct cheater detection cards, while ignoring certain cards that a logical SIA would choose, one needs to invoke cheater detection algorithms to explain *post*-interpretive inferences.

Most importantly, a careful application of RT_{Wason} to the background assumptions and pragmatic context for a trade leads to the conclusion that subjects should have answered either '*P & Q*' or '*P, Q, & not-Q*', which they did not. In other words, performance on selection tasks involving trades diverges from the predictions of relevance theory. This was true not only for Experiment 1, but also for other trades reported in the literature.

More specifically, the results for trades indicate that social contract algorithms are necessary for more than just the interpretive process. The world contains both the honest and the dishonest, and a relevance theoretic analysis of background assumptions and cognitive effects predicts that subjects should look for those who are honest as assiduously as they look for cheaters. But they did not. If there are relevance effects for trades, cheater detection subroutines *short-circuited* them. This means that social contract subroutines need to be invoked not just in the interpretive process, but to explain inferences made after that process is complete.

Indeed, by invoking social contract algorithms in the interpretive process and their cheater detection subroutine in the post-interpretive card selection process, one arrives at a single, simple explanation for performance on all rules involving trades, including the perspective shift experiments. The same explanation covers social laws that have the form of social contracts as well, and elegantly explains why social laws that are not social contracts (and not from other adaptively significant domains, such as precautions) fail to elicit high levels of performance (for examples, see Cosmides, 1989; Cosmides & Tooby, 1992).

For social contract laws, relevance theory and social contract theory may perhaps lead to the same prediction (although see Section 5). But for selection tasks involving trades, the combination of RT_{Wason} and social contract algorithms leads to worse predictive validity than invoking social contract algorithms alone. On grounds of

generality, then, one should prefer social contract theory to RT_{Wason} for all selection tasks involving social contracts. In short, it appears that when social contract effects and relevance effects lead subjects in diverging directions, social contract algorithms override relevance effects. This is just what an ecological rationality perspective predicts: when two or more inference engines are activated, the inference engine with the richer, more informative set of inferences should dominate reasoning outcomes.

We welcome relevance theory's application to the Wason task as potentially offering new explanatory insights in certain content areas, such as ordinary descriptive rules and denials. But to provide a truly *general, theoretically exclusive* account of selection task performance, relevance theory would have to invoke only logical SIAs, which are content-general. The analysis above shows that logical SIAs are not sufficient. Indeed, as we will show, they are not even necessary.

3. Experiment 2: are denials necessary?

According to SCG's analysis, one should not find high levels of '*P & not-Q*' answers on a selection task unless the conditional is interpreted as a denial (or, in the case of deontic rules, as a prohibition). In Experiment 2 we tested this claim by placing the conditional in a pragmatic context that blocks a denial interpretation. At the same time, we varied the testing strategy in a way that should make no difference to RT_{Wason}, but that does make a difference to hazard management theory, to see which theory successfully predicts performance.

3.1. Using pragmatics to block inference (c)

Pragmatically, denying something makes sense only if the denier thinks others have a prior expectation that it might be true (Wason, 1965). So by removing any prior expectation that *P-and-(not-Q)* is true (or might occur), one can prevent a conditional from being interpreted as a denial. This should prevent inference (c) from being made, and thereby prevent subjects from selecting the *not-Q* card.

When you are sincerely puzzled about what is true or what might happen, and ask someone to enlighten you, you are, from a pragmatic point of view, announcing that you have few prior expectations about these issues. A reply to such a question would therefore be difficult to interpret as a denial. The most pragmatically felicitous interpretation of such a reply is that it is a sincere attempt to answer the question. Hence, a simple manipulation to block a denial interpretation is to make the conditional statement the reply to a question. This is the pragmatic device we used in Experiment 2.

Assume, for example, that you had never seen or heard of an oven mitt. You then see extravagantly-colored, quilted, over-sized mittens hanging in a store. When you ask what they are for, the clerk might reply, 'If you take a pan out of the oven, then you wear these mitts to avoid being burned.' The clerk is not denying your prior expectation – you had none. He is simply offering the requested information. The most relevant logical form that could be assigned to a conditional rule in this case

would be $\forall x\ (Px \rightarrow Qx)$ (e.g. any time you take a pan out of the oven, then wear these mitts to avoid getting burned). The most relevant deductive consequence would be inference (a): that if an event has feature *P*, it also has feature *Q*. RT_{Wason} predicts that subjects test the truth of a rule through those of its deductive consequences that they spontaneously infer. So subjects asked whether the rule is true should choose the *P* card alone; if they also make inference (b), they might choose the *Q* card as well. But because the pragmatic context blocks inference (c) – the clerk was not denying a prior expectation – subjects should fail to choose the *not-Q* card.

3.2. Varying the testing strategy

RT_{Wason} proposes that people use content-general testing strategies, such as verification or falsification. If inference (c) has been made, then instructions to look for violations may increase the relevance of the *not-Q* card, compared to instructions to see if the rule is true. But if inference (c) has not been made, then subjects will fail to realize that *not-*[*P-and-(not-Q)*] is a deductive consequence of the rule. Since they won't have computed what counts as a violation, instructions to look for violations will be ineffective. For this reason, RT_{Wason} predicts that verification versus falsification instructions will not affect performance on rules that are not interpreted as denials. Indeed, that is exactly how SCG explain an otherwise puzzling phenomenon: the lack of instructional effects on ordinary descriptive rules.

According to RT_{Wason}, people choose either *P* alone, or *P* and *Q*, for ordinary descriptive rules because these are assigned the same logical form as the conditional offered in reply to the oven mitt question: $\forall x\ (Px \rightarrow Qx)$. Such rules trigger the deductive consequence in inference (a) and, sometimes, that in inference (b). Under these conditions, subjects fail to choose the *not-Q* card. Moreover, this failure is very robust. As SCG (p. 42) point out, it occurs not only when subjects are asked 'to determine whether the rule is true', but also when they are asked 'to see if *there are any cases that violate* the rule' (as well as many other variations). This is quite remarkable: after all, one might reasonably expect the instruction to look for cases that violate the rule to lead straightforwardly to choosing the *not-Q* card, especially when compared to instructions asking subjects to determine the rule's truth (a different and more complex question). Yet performance on abstract and descriptive versions of the selection task is not improved by violation instructions alone (Chrostowski & Griggs, 1985; Griggs, 1984; Jackson & Griggs, 1990; Kroger, Cheng & Holyoak, 1993; Manktelow & Evans, 1979; Reich & Ruth, 1982; Valentine, 1985; Yachanin, 1986, Experiment 2; although see Griggs, 1989; Platt & Griggs, 1993, Experiment 3).

To fully appreciate how odd this is, consider the following: if you ask subjects whether a card with *P* on one side and *not-Q* on the other violates the rule, 'If *P* then *Q*', they know that it does, even when the content of the rule is arbitrary (Manktelow & Over, 1987). So they do understand what counts as a violation and, in the selection task, they are directly instructed to look for violations. Yet they still fail to choose the *not-Q* card.

This phenomenon is not puzzling, however, if RT_{Wason} is correct in saying that the

rule is tested only through those of its deductive consequences that are spontaneously inferred, and that inference (c) is not triggered by these rules. By this reasoning, violation instructions should not help whenever inference (c) is blocked. And making the conditional a reply to a sincerely asked question should do just that.

In fact, we have taken SCG's logic one step further. In one condition of Experiment 2, we used standard truth-seeking instructions (the *Standard* condition). But in the other condition, we did not ask subjects to look for violations. Instead, we asked them to determine whether anyone is endangering themselves (the *Precaution* condition). If RT_{Wason} predicts that instructions to look for violations should be insufficient to elicit '*P & not-Q*' answers in response to a universally quantified rule, then it predicts the same for endangerment instructions. After all, RT_{Wason} is based on the generation of *logical* equivalences, and being in danger is not a concept in logic. It is, however, a concept in hazard management (precaution) theory, derived from an evolutionary functional analysis of what conceptual primitives and procedures will be necessary to make spontaneous inferences about dangers and their avoidance.

In Experiment 2, we used a rule similar to the oven mitt one above. It provides information that could be useful in reducing hazards. To the extent that subjects accept that the precaution is effective, instructions to 'look for individuals endangering themselves' should engage the hazard management system's checking routine, causing subjects to choose the *engaged in hazardous activity* card (*P*) and the *did not take precaution* card (*not-Q*). In contrast, truth-seeking instructions suggest that the effectiveness of the precaution is not yet known. In this situation, we would not expect the checking routine to be engaged. As a result, most subjects should fail to answer '*P & not-Q*', for the same reasons that they fail to do so on any ordinary descriptive rule (indeed, we find much to agree with in SCG's explanation for low levels of '*P & not-Q*' on ordinary descriptive rules).

3.3. Method

3.3.1. Subjects

An additional 60 subjects, different from those in Experiment 1, participated in Experiment 2. They were randomly assigned to one of two groups: 30 to the *Standard* condition and 30 to the *Precaution* condition.

3.3.2. Procedure and materials

The procedure was the same as for Experiment 1. The second page of each problem booklet contained a Wason selection task describing an unfamiliar scenario in which a returning tribesman sees some bright orange jackets, can't figure out why they are there, and so asks a fellow tribesman 'What are these for?' (see Fig. 3). The reply was in the form of a conditional: 'If you go hunting, then you wear these jackets to avoid being shot'. A question-and-answer context was used so that the conditional would not be interpreted as a denial. This part of the problem was identical for both the *Standard* and the *Precaution* conditions. Note that the conditional simply informs the asker about the function of the jackets. It is not presented

as a deontic rule prescribing how one ought to behave. Nowhere is it said or implied that one is *obligated* to wear these jackets when hunting, or that one is *permitted* to hunt only when wearing a jacket, or that one has violated any social law by failing to wear a jacket when hunting. Because the speaker's intention is to provide information about the jacket's function, his statement cannot be interpreted as *forbidding* people to hunt without jackets or as *denying* that people hunt without jackets.

Subjects received one of two versions of the selection task. In the *Standard* version, the tribesman was described as being uncertain about whether the conditional stated is true. In the *Precaution* version, the tribesman was described as being uncertain about whether all of his fellow tribesmen 'know about the jackets' and 'are needlessly endangering themselves'. Finally, in the *Standard* version, the condi-

You are a Kalama tribesman. While you were away on a hunting trip some anthropologists visited your village. The anthropologists often bring gifts for your tribe and this time you notice that they brought and left several bright orange jackets. You can't quite figure out why the anthropologists brought so many of the same jackets so you ask one of your fellow tribesmen **"What are these for?"** He tells you **"If you go hunting, then you wear these jackets to avoid being shot".**

Standard Version	**Precaution Version**
You are not sure if what he said is true so	You think the jackets are a great idea, but you are concerned that some of the other tribesmen might not know about the jackets and are needlessly endangering themselves.

You decide to watch what some of them do. The cards below represent four tribe members that you watched. Each card represents one person. One side of the card tells whether or not the person went hunting, and the other side of the card tells whether or not that person wore an orange jacket.

Standard Version	**Precaution Version**
Indicate only those card(s) you definitely need to turn over **to see if what your fellow tribesman said ("If you go hunting, then you wear these jackets to avoid being shot") is true.**	Indicate only those card(s) you definitely need to turn over **to see if any of these people are endangering themselves.**

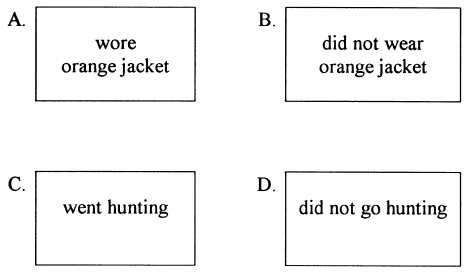

Fig. 3. Selection tasks used in Experiment 2. The only differences between the *Standard* and *Precaution* versions of the task are indicated in the boxes.

tional is restated and subjects are asked to indicate which cards they would need to turn over to see whether the conditional is true. This parallels the instructions given to subjects in the problems employed by SCG. In the *Precaution* version, subjects are asked to indicate which cards they would need to turn over to see if any of the tribesmen are endangering their lives.

3.4. Predictions

3.4.1. Relevance theory

By making the conditional an answer to a question in both versions, we have blocked the denial interpretation: the questioner asked what the jackets were for, having no prior expectations about what the answer would be. In this context, the logical form assigned to the rule should be $\forall x\ (Px \rightarrow Qx)$ which, according to SCG, should lead to low levels of '*P & not-Q*' answers in both versions. Instructions to look for verification versus falsification should have no effect, for the reasons discussed above. Moreover, these logical testing strategies are the only ones available to RT_{Wason}. To the extent that the *Precaution* version asks about a non-logical issue – is a tribesman endangering himself? – then this is outside the scope of SCG's theory. Thus, RT_{Wason} predicts low levels of '*P & not-Q*' responses for both the *Standard* and *Precaution* versions.

Note that in both versions, instances of *P-and-(not-Q)* are identical. These are cases in which a person went hunting without wearing an orange jacket. Thus, the cognitive effort of representing *P-and-(not-Q)* must be the same in both versions. The cognitive effect of finding such cases should be similar too: regardless of the final question, these are cases in which a person might be in danger (if the jackets have the protective effect claimed for them).

3.4.2. Hazard management theory

The fact that the conditional is the answer to a question is irrelevant to the hypothesis that there are content-dependent computational mechanisms that are functionally specialized for reasoning about hazards and precautions. By hypothesis, the function of these mechanisms is to check to see whether a person (or other entity) is in danger by virtue of having failed to take appropriate precautions in a hazardous situation (more generally, in a situation that could have negative consequences on a person or other valued entity). In the *Precaution* version, where subjects are explicitly asked to see if any tribesmen 'are endangering themselves', these mechanisms should cause subjects who believe that the jackets will have the protective effect to look for cases in which tribesmen may have engaged in the hazardous activity (the 'went hunting' card – *P*) without taking the precaution (the 'did not wear orange jacket' card – *not-Q*). Thus, the *Precaution* version should elicit high levels of '*P & not-Q*' responses. In contrast, the *Standard* version does not ask subjects to see who might be endangering themselves. Instead, it asks what information you would need to decide whether the claim about the practice of wearing jackets while hunting is even true. As such, it is like any other descriptive rule, and should elicit the response typical of descriptive rules that are universally quantified: low levels of '*P & not-Q*'.

It should not engage the hazard management mechanism's checking routines, because these are designed to operate when one already knows that a particular precaution is effective.

3.5. Results and discussion

The full pattern of card selections is given in Table 5. It shows that, as predicted by hazard management theory, the percentage of subjects who answered '*P & not-Q*' was substantially higher on the *Precaution* version (50%) than on the *Standard* version (17%) (50% versus 17%: test of proportions, $Z = 2.74$, $P < 0.004$).

As can be seen in Table 5, the predominant responses on the *Standard* version of the task were '*P*' (27%) and '*P & Q*' (23%). According to SCG's interpretation of the selection task, these are the patterns of selection expected if subjects assigned the conditional the logical forms $\forall x\ (Px \rightarrow Qx)$ and/or $\exists x\ (Px\ \&\ Qx)$, respectively. The predominant response on the *Precaution* version was '*P & not-Q*' (50%) followed by the *P* card alone (20%). In the *Precaution* version, the majority of subjects chose cards consistent with the desire to see who was endangering themselves, even though no deontic rule or social law was stated, and the conditional was not issued to deny a prior claim or expectation.

3.5.1. Implications

These results show several things:

(1) The pragmatic context of denial is not necessary to elicit high levels of '*P & not-Q*' responses. In both versions, the rule was descriptive, and placed in a pragmatic context in which it would be most sensibly interpreted as a universally quantified statement (inference (a)). Yet 70% of subjects chose the *not-Q* card in the *Precaution* version, and 50% chose all and only *P* and *not-Q*.

(2) A rule need not be deontic to elicit high levels of '*P & not-Q*' responses. This

Table 5
Experiment 2: pattern of card selections (percentage of subjects)

Pattern	Condition	
	Standard	*Precaution*
P, not-Q	16.7	50.0
P	26.7	20.0
P, Q	23.3	6.7
P, Q, not-Q	10.0	3.3
not-P, not-Q	0.0	10.0
not-P, Q	6.7	0.0
not-Q	6.7	0.0
Q, not-Q	0.0	6.7
None	6.7	0.0
not-P	3.3	0.0
P, not-P, Q	0.0	3.3

result supports hazard management theory, and poses a serious problem for those theories that claim this response is elicited only by deontic rules (e.g. Manktelow & Over, 1995). Furthermore, subjects were not asked to look for violations of the rule: they were asked to see who might be in danger. 'Danger' is not a concept in a deontic logic, just as it is not a concept in the predicate calculus.

(3) Contrary to the predictions of RT_{Wason}, the *Standard* and *Precaution* versions elicited different response profiles. To be consistent, RT_{Wason} could accommodate this result only by positing differences between the two conditions in the ease with which cases of *P-and-(not-Q)* can be represented, or differences in their cognitive effects, such that finding cases of *P-and-(not-Q)* would cause more belief revisions in the *Precaution* version than in the *Standard* one. But there are no differences in cognitive effort: *P-and-(not-Q)* represented exactly the same state of affairs in both versions (hunting with no jacket). Moreover, given that the two versions are designed to parallel each other in so many respects, it seems unlikely that finding cases of *P-and-(not-Q)* would cause more belief revisions in one condition than the other. But the key point is that if it did, it would have to be in the *Standard* version. After all, in both versions the same people hunting without orange jackets are in danger, assuming the jackets have the protective effect claimed for them. So when it comes to consequences for the individuals hunting without jackets, the effects are the same in both versions. But finding people hunting without jackets in the *Standard* version allows more extensive belief revisions. Remember, the *Standard* version does not raise the possibility that others are ignorant of the jacket's function. So if hunters are not wearing the jackets, this suggests that (i) they do not believe the claim that the jackets have a protective effect, (ii) the person who told you they are useful is either unreliable or lying to you, or (iii) there is some additional, interesting feature of the situation which has yet to be explained. In short, even if there are more cognitive effects on one version than another before RT_{Wason}'s stopping rule is applied, more cognitive effects would be triggered by the *Standard* version.

We would not want to press the analysis of cognitive effects too hard – and we would hope that SCG would agree with us. RT_{Wason} gives clear predictions about cognitive effects when the background assumptions are clear, as in denials, the drinking age problem, or a trade. However, the theoretical variable of cognitive effects, when used explanatorily outside of certain well-specified contexts, can easily become a free parameter that makes RT_{Wason} predictively weak but retrodictively protean. For example, if one were to extend the notion of cognitive effects to encompass any state of affairs that strikes one person or another as more 'interesting', RT would quickly become unfalsifiable. One can often reason backward from a result – or worse, from one's own reactions on reading a problem – to the conclusion that one state of affairs *must* have had more cognitive effects than another. But to have predictive bite, one must be able to reason forward from a theory, not just backward from results. This requires precise theories about just what conditions produce cognitive effects. Without such a specification as part of the theory, one is relying on intuition rather than explaining its cognitive basis (see Section 5).

(4) The different response profiles elicited by the *Standard* and *Precaution*

versions are those predicted by hazard management theory. Moreover, the factors in that theory that lead to this prediction fall outside the scope of RT_{Wason}. First, hazard management theory does not hinge on a rule being interpreted as deontic. A conditional that informs one about risks that can potentially be circumvented may provide useful information, but it need not be turned into a social rule that one is then obligated to follow. Indeed, one may often have good practical or even moral reasons for not following such a derived rule. Second, although the rule in Experiment 2 was a descriptive one, its content was relevant to managing a hazard. What it *described* was the protective function of orange jackets: that they can reduce the hazards of hunting. Thus, it can be mapped onto the content-dependent representation of a precaution rule. Third, high levels of '*P & not-Q*' responses were elicited only when the subject was cued to accept the efficacy of the jackets, and then asked to look for individuals who might be in danger: again, conditions that should activate the checking routine of hazard management algorithms.

Note that the final question in the *Precaution* version was non-logical. This is important, because the joint claims that RT_{Wason} 'explains the selection task' and that the selection task cannot be used to study content-specific inference mechanisms hold only insofar as RT_{Wason} can explain results *without invoking content-specific inference mechanisms*. But the only inference tools available to RT_{Wason} for this purpose are content-independent logical SIAs. RT_{Wason} alone cannot explain high levels of '*P & not-Q*' answers in response to a question about danger. To account for this result, a content-specific concept and checking routine would have to be added to RT's arsenal of SIAs. But if these are needed to explain performance, then, obviously, selection task results do reveal properties of the computational machinery that guides people's selections.

3.5.2. Do the instructions for the precaution version highlight violations?

According to RT_{Wason}, instructions can affect the relevance of content-independent testing strategies, such as verification and falsification. Could this have made a difference in Experiment 2, given that the same logical form should have been assigned to the rule in both conditions? We think the answer is a clear no. Regardless of instructions, subjects should look for violations when inference (c) is made, but not otherwise (see above). But there are no background assumptions that would lead to inference (c) being made for the *Precaution* version that would not apply equally to the *Standard* version (see Appendix B).

There is one other difference between the *Standard* and the *Precaution* versions. In the *Standard* version, the subject is asked to reason *about* the rule (the purpose of turning over cards is to determine whether the rule holds), whereas in the *Precaution* version, the subject is asked to reason *from* the rule (assuming the rule holds, the purpose of turning over the cards is to see whether anyone is in danger). This is an important difference for certain deontic theories (e.g. Manktelow & Over, 1990, 1991, 1992), because it means there is at least one sense in which the tasks are not logically equivalent. It is not, however, an important difference for RT_{Wason}. In the interests of providing a general account of the selection task, SCG (pp. 83–84) specifically reject this as an explanation for differences in performance. Curiously,

this rejection occurs in the context of a brief discussion of a key experiment by Gigerenzer and Hug (1992) – but without then providing a relevance theoretic explanation for the results of that experiment.

Although not designed to test against RT_{Wason}, the logic of Gigerenzer and Hug's experiment is similar to the logic of Experiment 2, their results are parallel to our results, and one can draw similar conclusions from it with respect to RT_{Wason}. The main difference is that their experiment was designed to test predictions of social contract theory, whereas Experiment 2 was designed to test predictions of hazard management theory. Both experiments indicate that content-specific inference systems (social contract algorithms, hazard management algorithms) organize reasoning and override relevance effects in their respective domains.

3.5.3. A parallel case with social contracts

Gigerenzer and Hug (1992) argued that what is important for improved performance on social contract versions of the selection task is not merely that the rule be interpreted as a social contract, but that the task be to detect cheaters, i.e. individuals illicitly taking the benefit specified in the social contract rule. They tested this by comparing performance on selection tasks using identical social contract rules. In the *cheating* version, looking for violations was the same as looking for cheaters; in the *no cheating* version, the subject was asked to look for violations to determine whether the social contract rule is in effect. They conducted multiple tests of this hypothesis, using several different social contract rules. We will illustrate using their cabin problem.

In the *cheating* version, the Swiss Alpine Club has made a rule for use of their cabin by hikers: 'If one stays overnight in the cabin, then one must bring a load of firewood up from the valley.' In this version, looking for violations was equivalent to looking for individuals who have illicitly benefited themselves, i.e. who have cheated. In the *no cheating* version, a German hiking in the Swiss Alps sees people bringing loads of firewood into a Swiss Alpine Club cabin, and wonders why. His friend suggests that the Swiss may have the same social contract rule as the Germans, that is, 'If one stays overnight in the cabin, then one must bring a load of firewood up from the valley.' Another possibility is also suggested: that wood is supplied by local club members who do not use the cabin. To settle the question, the subject was asked *to assume that the proposed social contract rule is in effect*, and then look for violations of it. Note that the intent here was not to catch cheaters. In this case, violations of the proposed social contract rule could occur simply because the Swiss Alpine Club never made such a rule in the first place.

The result: almost twice as many subjects answered '*P & not-Q*' for the *cheating* as for the *no cheating* versions, and the *no cheating* versions elicited about the same level of '*P & not-Q*' responses as descriptive rules. Yet in both versions, (i) the rule in question was a social contract rule, (ii) the subject was asked to assume that it was in effect, and (iii) the subject was asked to look for violations. The only difference was whether looking for violations was framed as a means of detecting cheaters or as a means of determining whether the rule actually was in effect. This distinction is not relevant to RT_{Wason}.

This is a very difficult result for RT_{Wason} to explain. What is important to RT_{Wason} is the interpretation of the rule: here, that the rule is seen as deontic and as forbidding cases of *P-and-(not-Q)*. This was clearly the case in both versions. In the *no cheating* version, when the friend suggested that the Swiss might have the same social contract rule as the Germans, he was claiming that the Swiss might have a deontic rule forbidding people to stay in the cabin without having brought a bundle of wood. So when subjects were asked to assume the rule is in effect, and then look for violations of it, the majority of them should have chosen the 'stayed in cabin' card and the 'did not bring wood' card, i.e. *P* and *not-Q*. Yet they did not.

One might be tempted to say that when finding a violation means you have found a cheater, this has more 'cognitive effects' than in the *no cheating* version. This is a temptation that should be resisted by advocates of any content-general account, since they will want to steer clear of making cognitive effects a retrodictive free parameter. If 'cognitive effect' refers to belief revision and the making of further inferences, as SCG propose, then surely discovering whether the Swiss do, or do not, have the social contract rule in question is going to lead to at least as many inferences and revised beliefs: this discovery will determine whether the hikers are going to have to go down the mountain to get wood, or stay at high altitude and sleep elsewhere (and perhaps freeze!), or use the cabin illicitly, for example. We see no principled way of claiming that this involves *fewer* inferences (or, indeed, less important consequences for the hikers' health and safety) than finding a cheater or two in the other version. It will not do to say that finding cheaters is more 'salient' or more 'interesting' than finding other kinds of violations: this just re-describes the results. To preserve both its generality and its testability, RT_{Wason} would have to explain differential salience in a theoretically principled way. That means without invoking content-specialized mechanisms, such as social contract algorithms, and using only the content-general theoretical tools at its disposal, such as conversational pragmatics and logical SIAs. We do not think this can be done.

SCG's attempt to subsume deontic versions of the selection task under RT_{Wason} depends on deontic rules activating the same cognitive processes as descriptive rules, so that "in spite of their logical differences, deontic and descriptive versions are not psychologically so different after all" (pp. 83–84). But if deontic and descriptive rules evoke the same cognitive processes, then surely identical deontic rules do too. This leaves RT_{Wason} with no explanation for the fact that (i) the same deontic rule elicited different levels of '*P & not-Q*' responses in the *cheating* and *no cheating* versions, and (ii) deontic rules that should have been assigned a 'forbid' interpretation (the *no cheating* versions) elicited relatively low levels of '*P & not-Q*' responses, just like the ordinary descriptive (i.e. non-denial) problems that Gigerenzer and Hug (1992) tested.

In contrast, Gigerenzer and Hug (1992) argue that quite different cognitive processes are evoked, not just by descriptive versus deontic problems, but by identical social contract problems, as a function of whether the instructions activate a specific post-interpretive inference mechanism: a cheater detection subroutine. They predicted the difference between the *cheating* and *no cheating* versions in advance, and this prediction was based on the hypothesis that (i) a functionally-specialized

checking routine – the cheater detection procedure – is applied to the already interpreted social contract rule, and (ii) this routine embodies a content-dependent definition of violation – *cheating* – that is specific to the domain of social contracts.

We shall explore the implications of content-dependent definitions of violation further in Experiment 3.

4. Experiment 3: one notion of violation, or many?

Verification and falsification are testing strategies, derived from logic, that are content-independent. These are the only testing strategies available to RT_{Wason}, if it is to explain selection task performance without recourse to content-specialized reasoning mechanisms. In Experiment 3, we show that they are insufficient to explain selection task performance on social contracts and precaution rules.

Given its various theoretical commitments, RT_{Wason} could not handle the results of the Gigerenzer and Hug (1992) cheating/no cheating experiment. It is often possible to create alternative explanations for a single result. But to explain the full array of results on selection tasks involving social contract rules, we believe a testing strategy that involves the search for *cheaters* – individuals who have illicitly taken the benefit specified in a social contract rule – must be invoked (e.g. Cosmides & Tooby, 1992, 1997). Precaution rules have not been tested as extensively as social contract rules, but we predict that a similar conclusion will turn out to be warranted: that these activate checking routines that embody content-specialized testing strategies and definitions of violation, which are sensitive to conditions of endangerment (Cosmides & Tooby, 1997; Fiddick, 1998; Rutherford et al., 1996).

The issue dividing RT_{Wason} from these other theories is not whether testing strategies per se make a difference. According to social contract theory, instructions to 'look for cheaters' should indeed elicit different answers than (say) truth-seeking instructions. But the content-specific theories propose that there are domain-specific, content-dependent testing strategies. Looking for cheaters (social contract theory) or looking for people in danger (hazard management theory) requires the search for very specific, content-defined categories of information, and the mechanisms that accomplish this must be able to operate in real life situations – regardless of the surface form that an explicit rule may take, and even in cases where the rule is implicit. These computational requirements cannot be met by a content-independent testing strategy, such as looking for a true antecedent and a false consequent (see Cosmides, 1989; Cosmides & Tooby, 1992, and discussion below).

4.1. The problem with logic

To be truly content-general, RT_{Wason} must rely solely on logical SIAs. It must posit that when subjects are asked to 'look for violations', they look for *logical* violations, that is, they apply the content-independent definition of violation from logic (true antecedent and false consequent). When subjects assign the conditional rule the logical form specified in inferences (a) and/or (c), there is only one conjunction of features that can violate it: *P-and-(not-Q)*. The predicate calculus admits no other

way to falsify these logical forms. Hence, if 'look for violators' instructions lead subjects to apply the content-independent definition of violation drawn from logic, then RT_{Wason} predicts that they should select the *P* and *not-Q* cards, regardless of content. One straightforward way to rule out the hypothesis that subjects merely adopt a content-general falsificationist strategy on the selection task is to show that subjects cued to look for *different* violations make different card selections. This is the strategy we pursued in Experiment 3.

4.2. Different kinds of violations

The ecological rationality approach leads to the expectation that the mind makes content-sensitive distinctions other than simply 'true' and 'false'. This can be tested by creating rules that are susceptible to dual interpretations, with conflicting definitions of violation.

By choosing the content carefully, one can create a rule of the form *If P then Q* that can be interpreted as either a social contract or as a precaution rule. The drinking age problem is one example; the cholera problem (Cheng & Holyoak, 1985) is another (see SCG, p. 84, and Cosmides, 1989, for a social contract interpretation). Both rules can be assigned one of two representations: (i) *If you accept the benefit, then you must satisfy the requirement* (social contract) or (ii) *If you engage in a hazardous activity, then take the appropriate precaution* (precaution rule). There are certainly ways to discover which representation an individual subject has assigned to one of these rules. But the drinking age and cholera problems cannot illuminate the issue of content-sensitive definitions of violation because for these particular rules, looking for cheaters and looking for endangerment would lead subjects to choose the same cards: *P* and *not-Q*.

It is possible, however, to create a rule that can be assigned either interpretation, but where cheater detection would lead to a different pattern of card choices than looking for individuals who are in danger. For example, the conditional rule, *If you spray insecticide on my garden, then I will give you a gas mask to use and keep*, is of the form: *If Hazardous Requirement, then Precautionary Benefit*. If said to an avid gardener, it could be interpreted as either a social contract (If you agree to satisfy the requirement, then I will provide the benefit) or as a precaution (If you engage in a hazardous activity, then take a precaution). A cheater would be someone accepting the benefit without satisfying the requirement: someone who took the gas mask, but didn't spray the insecticide, i.e. *Q-and-(not-P)*. In contrast, a person would be endangered by engaging in the hazardous activity without taking the appropriate precaution: someone who sprayed the insecticide, but didn't use the gas mask, i.e. *P-and-(not-Q)*. Hence, social contract theory and hazard management theory, when considered together, predict that there is more than one way to violate this rule.

In Experiment 3, we gave subjects a deontic version of the selection task employing a rule of the form: *If Hazardous Requirement, then Precautionary Benefit*. Minimal changes were made in the problem context to trigger subjects to represent this rule as either a social contract or as a precaution and to instruct subjects to look for cheaters or for people in danger, respectively. The domain-specific view treats

these as two separate violations and, therefore, predicts that they should elicit a different pattern of card selections. In contrast, RT_{Wason}, because of its dependence on a content-free, logical definition of violation, predicts that subjects will try to falsify the rule by picking the *P* and *not-Q* cards in both conditions.

4.3. Method

4.3.1. Subjects

An additional 120 subjects participated in Experiment 3. They were randomly assigned to one of two groups: 60 to the *Privilege* condition and 60 to the *Risk* condition.

4.3.2. Procedure and materials

As before, the first page contained general instructions. The second page contained a Wason selection task employing the rule: *If you make poison darts, then you may use the rubber gloves*. This rule was embedded in a scenario in which an anthropologist has brought a limited supply of gloves for tribespeople to use while engaging in the hazardous activity of making poison darts (see Fig. 4). The gloves, it was claimed, offered protection against the poison, but, because the supply was limited, their use was restricted to people making poison darts. In this context, the rule was of the form: *If Hazardous Requirement, then Precautionary Benefit*. This means it can be assigned two alternative abstract representations: that of a precaution rule (If you engage in a hazardous activity, then take the precaution) or of a social contract (If you satisfy the requirement, then you [are entitled to] take the benefit). (Bracketed words represent concepts that the domain-specific theories predict subjects will import when interpreting the rule.) In other words, the same action – *making poison darts* – would be assigned the abstract representation 'Hazard' under the precaution interpretation and 'Requirement' under the social contract interpretation. Similarly, *using rubber gloves* would be assigned the abstract representation 'Precaution' under the precaution interpretation and 'Benefit' under the social contract interpretation.

Subjects received one of two versions of the selection task. In the *Privilege* version, the anthropologist is concerned that the tribespeople are abusing the privilege of wearing the gloves. In the *Risk* version, the anthropologist is concerned that some of the tribespeople are risking their lives. As one can see in Fig. 4, these manipulations were achieved with minimal changes in the wording of the problems. The purpose of varying the last question was to change the nature of the 'violation' that subjects would be inclined to look for and, therefore, flip people into just one of the two alternative representations of the rule. Looking for people who are 'abusing the privilege' should lock in the social contract representation, whereas looking for people who are 'risking their lives' should lock in a precaution representation.

4.4. Predictions

4.4.1. Social contract theory and hazard management theory

Neither problem explicitly asks the subject to 'look for violations'. However, just

You are an anthropologist studying the Kalama tribe.

The Kalama hunt with blowguns and poison darts. The poison is a powerful neurotoxin obtained from a small tree frog and has been known to kill humans too. In fact, several Kalama have died preparing poisoned darts when the poison got onto their exposed skin. Last time you visited the Kalama, you observed this happen, so you brought a supply of rubber gloves for the Kalama tribesmen to wear to avoid contact with the poison when making darts. You were only able to bring a limited supply of rubber gloves so you restricted their use to when someone's life might be in danger. You told the tribesman:

"If you make poison darts, then you may use the rubber gloves."

You are concerned that members of the Kalama tribe are

Social Contract Version	Precaution Version
abusing this privilege,	risking their lives,

...so yesterday you watched what some of them did. The cards below represent four tribe members that you watched yesterday. Each card represents one person. One side of the card tells whether or not the person made poison darts, and the other side of the card tells whether or not that person wore rubber gloves.

Indicate only those card(s) you definitely need to turn over **to see if any of these people are**

Social Contract Version	Precaution Version
abusing the privilege.	**risking their lives.**

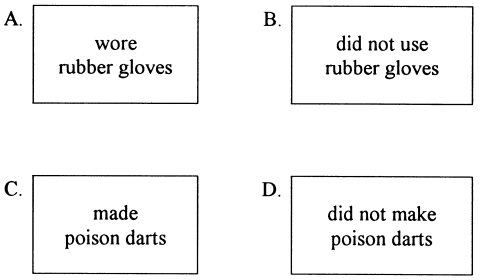

Fig. 4. Selection tasks used in Experiment 3. The only differences between the *Privilege* and *Risk* versions of the task are indicated in the boxes.

as the instruction to 'look for cheaters' is interpreted as an instruction to look for violations of a specific kind, so should the instruction to look for people 'abusing the privilege' or for people 'risking their lives'.

4.4.1.1. Privilege condition The instruction to see if anyone is 'abusing the privilege' should activate a social contract interpretation of the rule, because social contracts are conditionals that regulate privileges (benefits, access to which has been restricted in some way). According to social contract theory, cheating is defined as taking a benefit without meeting the requirement that provision of that benefit was made contingent upon. In this situation, using the gloves is the benefit (it is a benefit because it provides protection), and provision of that benefit is restricted

to times when one is making poison darts (a restriction that helps preserve the lifespan of a very limited supply of gloves). So a person 'abusing the privilege' would be someone who is illicitly accessing the benefit: using the gloves when they are doing something other than making poison darts. To find such individuals, one must choose the *Q* card – 'wore rubber gloves' – and the *not-P* card – 'did not make poison darts'. Thus, the *Privilege* condition should elicit '*not-P & Q*' from the majority of subjects. This is not a logically correct answer. Moreover, not under any of RT_{Wason}'s three interpretations of a conditional (a, b or c) does RT_{Wason} predict subjects to choose *not-P & Q*, *so the predictions between the two theories are here maximally discrepant.*

4.4.1.2. Risk condition The instruction to see if anyone is 'risking their lives' should activate a precaution representation. Risking one's life is engaging in a hazardous activity without taking appropriate precautions. In this situation, engaging in the potentially lethal activity of making poison darts is the hazardous activity, and wearing the rubber gloves is supposed to protect you from being poisoned while engaging in that hazardous activity. So a person risking his life would be someone making poison darts without wearing the rubber gloves for protection. To find such individuals, one must choose the *P* card – 'made poison darts' – and the *not-Q* card – 'did not use rubber gloves'. Thus, the *Risk* condition should elicit '*P & not-Q*' from the majority of subjects. The fact that this is the logically correct answer is a side-effect: by hypothesis, these cards are selected not because they are logically correct, but because hazard management mechanisms are designed to look for people at risk.

4.4.2. Relevance theory

In contrast, for RT_{Wason} the abstract representation of a deontic rule should be the same, whether the rule is a social contract or a precaution. Moreover, if RT is to explain the selection task, then it must do so *without invoking social contract or hazard management algorithms*. In other words, a truly general form of RT cannot claim that reference to privileges versus risks causes different logical forms to be assigned to this rule in the two conditions, because these are conceptual primitives of social contract and hazard management algorithms (respectively), and not of the predicate calculus. Thus, the rule must be assigned a logical form in accordance with the discourse structure and the rules of logic alone. This form will either be $\forall x (Px \rightarrow Qx)$ or *Forbid(P-and-(not-Q))*.[26] Whichever form is assigned, it should be the same for the *Privilege* and the *Risk* conditions, as the discourse structure is the same in both conditions (neither involves denials and the rule is, in both cases, described as deontic). Both the *Risk* and the *Privilege* versions imply that people may be violating the rule, thus the instructions make a falsification testing strategy more

[26] Note that the temptation to retranslate the rule as 'If you use rubber gloves, then you must make poison darts' is warranted only by social contract algorithms and is, therefore, not a move allowable by RT. Even if one were to retranslate the rule in this way, according to RT it should be done in both conditions. Hence, one would still predict that there would be no difference in responses between the two conditions (it would be 'wore rubber gloves' and 'did not make poison darts', i.e. '*Q & not-P*' when these values are defined with respect to the rule as given in the original problem).

relevant than a verification one. However, the only concept of violation available to a discourse comprehension module equipped only with logical inferences is the logical notion of violation: true antecedent plus false consequent. Hence, subjects should answer '*P & not-Q*' in both conditions. (If the 'may' in the rule were interpreted as indicating logical possibility, then it could be argued that the logically correct answer to both would be to choose no cards at all.)

4.5. Results and discussion

RT_{Wason} predicts that '*P & not-Q*' will be the predominant response in both conditions, but it was not. Although 70% of subjects produced this response in the *Risk* condition, only 5% did so in the *Privilege* condition. Subjects made different patterns of card selections depending upon the type of violation they were asked to look for, just as the domain-specific theories predicted.

4.5.1. Privilege condition

A person who is making poison darts without wearing the rubber gloves – *P-and-(not-Q)* – may be at risk, but he is not abusing the privilege. Accordingly, only 5% of subjects answered '*P & not-Q*' when asked to look for people abusing the privilege, whereas 58% produced the full cheater detection response – '*not-P & Q*' – in the same condition. These are the 'did not make poison darts' (*requirement not satisfied*) and 'wore rubber gloves' (*benefit accepted*) cards, respectively.

4.5.2. Risk condition

The opposite pattern was observed in the *Risk* condition. A person wearing the rubber gloves without making poison darts – *(not-P)-and-Q* – may be abusing the privilege, but he is not at risk. Accordingly, no one answered '*not-P & Q*' when asked to look for individuals risking their lives, whereas 70% of subjects checked for people in danger in the same condition, by answering '*P & not-Q*'. These are the 'made poison darts' (*engaged in hazardous activity*) and 'did not use rubber gloves' (*did not take precaution*) cards, respectively.

The distinct patterns of card selections elicited by each condition can be seen in Table 6. The answer, '*not-P & Q*' was given significantly more often in the *Privilege* than in the *Risk* condition (58% versus 0%: $Z = 7.03$, $P < 1.1 \times 10^{-12}$), and the answer '*P & not-Q*' was given significantly more often in the *Risk* than in the *Privilege* condition (70% versus 5%: $Z = 7.35$, $P < 1.0 \times 10^{-13}$). A χ^2 test of the interaction (Risk/Privilege versus '*P & not-Q*'/'*not-P & Q*') was highly significant ($\chi^2(1, N = 80) = 68.77$, $P < 1.1 \times 10^{-16}$). This interaction can be seen clearly in Fig. 5.

4.5.3. Implications

The purpose of Experiment 3 was to see whether providing instructions that emphasize violation would make a domain-general falsificationist testing strategy relevant. This was clearly not the case. The *Risk* versus *Privilege* conditions induced subjects to look for two entirely different *kinds* of violations. Yet there is only one

Table 6
Experiment 3: pattern of card selections (percentage of subjects)

Pattern	Condition	
	Privilege	*Risk*
P, not-Q	5.0	70.0
not-P, Q	58.3	0.0
P, Q	8.3	1.7
P	8.3	8.3
Q	6.7	1.7
not-P	5.0	0.0
not-Q	1.7	10.0
P, Q, not-Q	1.7	3.3
Q, not-Q	0.0	3.3
All	1.7	0.0
P, not-P, Q	1.7	0.0
P, not-P	1.7	0.0
not-P, Q, not-Q	0.0	1.7

notion of violation available in the predicate calculus: true antecedent plus false consequent.

This result makes sense if one assumes that (i) a cheater detection subroutine is designed to activate and act on a social contract representation of the rule, whereas a precautionary checking routine is designed to activate and act on a hazard/precaution representation, (ii) these two checking routines are computationally different and content-specialized, and (iii) their outputs should dominate relevance effects, if such exist.

Note how sensitive performance on the selection task was to a change in testing-strategy, when this change tapped into ecologically rational, content-dependent

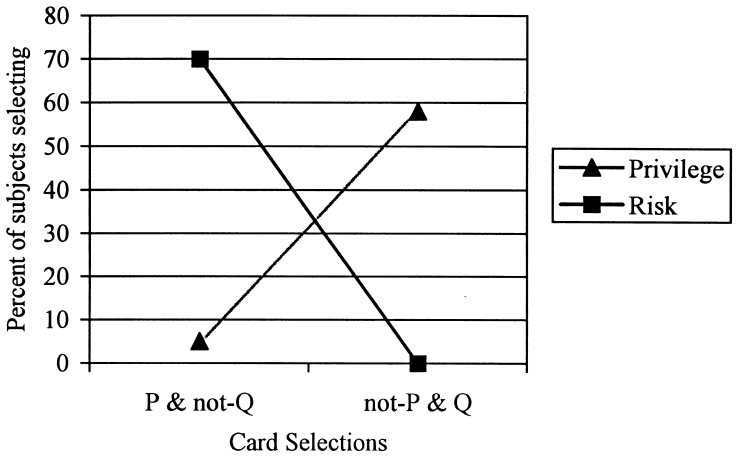

Fig. 5. Percentage of subjects answering '*P & not-Q*' or '*not-P & Q*' on the *Privilege* and *Risk* versions of the selection task used in Experiment 3.

notions of violation. This is quite different from what one finds for descriptive rules, where a large change in testing-strategy – from truth-seeking instructions to instructions to look for violations – causes little or no difference in performance.

4.5.4. Can relevance theory explain the choice of 'not-P & Q' in the privilege condition?

The fact that a majority of the subjects in the *Privilege* condition answered '*not-P & Q*' replicates previous studies employing 'switched' social contract rules (e.g. Cosmides, 1989; Gigerenzer & Hug, 1992). In her initial tests of social contract theory, Cosmides (1989) attempted to disambiguate the role of logical form and social contract form with the use of switched social contracts. A switched social contract is of the form: *If you have satisfied the requirement, then you are entitled to accept the benefit* (see Fig. 1). Cosmides argued that, if subjects were guided by the logical form of the rule and were searching for logical violations, then one would expect switched social contracts to elicit the same '*P & not-Q*' card selections as is observed on standard social contract versions of the task (especially when deontic operators, such as 'may', 'entitled', or 'must' are left out, as they were in her experiments). However, if a cheater detection routine operates on the social contract form of the rule, then subjects should continue to choose the cards corresponding to potential cheaters – the *requirement not satisfied* card and the *benefit accepted* card – even though these map onto the logical categories *not-P* and *Q* for a switched social contract. Subjects in the Cosmides (1989) experiments overwhelmingly answered '*not-P & Q*' – an illogical but adaptively sound answer – for switched social contracts, as predicted by social contract theory.

RT_{Wason} cannot readily explain performance on switched social contract problems without invoking social contract algorithms. SCG do not discuss the switched social contract experiments beyond the comment that the rules used were 'pragmatically awkward'. Awkwardness in communication is an insufficient explanation for why one regularly obtains a robust, specifically patterned set of results (*not-P & Q*) – card selections which were never observed previously, and which were not the predicted computational product of any reasoning theory other than social contract theory. To explain these results, one needs a theory of what meaning is reliably extracted by the subject from the putatively unclear utterance, and then what rules of inference are applied to that interpretation to arrive at the reliably obtained results. RT_{Wason} lacks such a theory.

In any case, while it could be argued that some switched social contract laws sound mildly infelicitous, this is manifestly not true for the great majority of switched personal exchanges, which are expressed in a pragmatic form that is commonly used in daily life, e.g. 'If you pay my moving expenses, then I will accept your professorship' (where the university is the potential cheater). The two personal exchanges tested in Cosmides, 1989 (Experiment 4) were completely natural (e.g. 'If I give you duiker meat, then you must give me your ostrich eggshell'). If the supposed awkwardness of the switched rules was intended by SCG to explain why subjects failed to make the logically correct choice on them, then the results of Experiment 3 pose a dilemma for RT_{Wason}. The rule employed in this experiment,

If you make poison darts, then you may use the rubber gloves, is a switched social contract. This leaves relevance theory with one of two contradictory outcomes: (i) the rule is pragmatically awkward – this might explain why subjects did not answer logically in the *Privilege* condition, but it fails to explain why subjects answered logically, choosing '*P & not-Q*' in the *Risk* condition; or (ii) the rule is not pragmatically awkward – this would explain why subjects made the logical '*P & not-Q*' selection in the *Risk* condition, but it fails to explain why subjects consistently made the illogical '*not-P & Q*' selection in the *Privilege* condition.

An alternative reading of SCG is that they believe their explanation for the choice of '*not-P & Q*' in the perspective shift experiments is also an adequate explanation for why people make this choice when faced with switched social contracts. But, as we showed above, to explain the prevalence of '*not-P & Q*' responses in the perspective shift experiments, *relevance theory must invoke social contract algorithms*. Social contract algorithms provided all the crucial inferences and the appropriate content-dependent definition of cheating. Consequently, a truly general form of RT_{Wason} – a form that does not rely on content-specific algorithms to explain the selection task – cannot explain the choice of '*not-P & Q*' on switched social contracts any more than it can explain this choice on social contracts involving perspective shifts.

5. General discussion

SCG argue that relevance $theory_{Wason}$ "explains the selection task", providing a "wholly general" account of performance, including all content effects. We think the evidence clearly contradicts this broad claim.

Any elaboration of relevance theory that incorporates and relies on social contract algorithms and other evolved specializations to make crucial inferences would not constitute a competing account of selection task performance. But RT_{Wason} is not an elaboration of this kind.

What gives RT_{Wason} the potential to be "wholly general" is its reliance on content-independent representations and procedures drawn from logic. This is what makes it a competing explanation for the intricate patterns of subject performance that have been advanced as evidence for domain-specific mental operations. However, so long as RT_{Wason} depends on and is restricted to representations of logical form and logical SIAs, it cannot explain performance on the selection tasks involving social contracts and hazard/precaution rules tested herein. The range of inferences licensed by logic is simply too constrained to account for the precisely patterned choices subjects actually make.

Indeed, taking a more synoptic view, the prior literature on social contract reasoning has predicted and confirmed a large range of content effects that are similarly beyond RT_{Wason}'s explanatory powers. RT_{Wason} cannot explain performance on social contracts that represent trades rather than social laws (Cosmides, 1989; and herein), ones with perspective shifts, ones where violations are not equivalent to cheating (Gigerenzer & Hug, 1992), or ones where the benefit and requirement terms are

switched. In fact, a careful application of the concepts most essential to RT_{Wason} – background assumptions, cognitive effects, cognitive effort, logical form and deductive consequences – leads to the conclusion that content-specialized social contract algorithms must be invoked to explain the patterns of response elicited by these problems (see Sections 2.3 and 4.5). In the case of trades, combining RT_{Wason} with social contract theory led to worse predictive validity than social contract theory alone.

Thus, RT_{Wason} cannot account for a substantial body of content effects involving social contracts and precaution rules. When the content of a problem activates one of these domains, performance tracks the predictions of social contract or hazard management theory, not RT_{Wason}. Indeed, if RT_{Wason} really explained the selection task for social contracts and hazard management problems, then performance on social contract and hazard management problems should change dramatically when elements central to RT_{Wason} – both logical and pragmatic – are varied. But it does not: logical connectives can be removed, denial contexts eliminated, and subjects asked to search for content-specific violations that do not correspond to logical violations. Yet, regardless of these manipulations, subjects steadfastly continue to search for cheaters and people in danger, just as the content-specific theories predict they will.

We would go even one step further. It is not clear that certain core claims of RT_{Wason} – such as that processing negations is effortful – are true across domains. For example, whereas processing explicit *nots* does seem to involve cognitive effort for descriptive rules (e.g. Wason & Johnson-Laird, 1972), it is not at all clear that this is true for deontic ones. Given a precaution rule such as 'If you walk in poison ivy, wear tall boots', in which case is it easier to tell that the requirement was not met: 'he did not wear tall boots' or 'he wore sneakers'? In the case of a trade, failure to live up to an agreement is most easily and unambiguously expressed by an explicit *not*: 'The villager did not give the farmer any corn'.[27] In pilot studies for the experiments reported in Stone et al. (1996, 2000), we found that neurologically impaired individuals were more likely to give correct answers to deontic rules when *not-Q* was expressed as an explicit, rather than implicit, *not*.

We are certainly not arguing that relevance theory in general is wrong or even that RT_{Wason} should be abandoned. Indeed, we are intrigued by certain aspects of RT_{Wason}, such as its account of performance on descriptive rules, both ordinary ones and those involving denials. In the long run, data may support RT_{Wason}'s explanation for performance in these delimited domains, and perhaps in others yet to be identified.

But regardless of whether this happens, RT_{Wason} does not explain performance for domains in which there exist powerful reasoning specializations that involve extralogical procedures. The data reported herein, as well as the literature discussed, show that the inference mechanisms activated by social contracts and precaution rules short-circuit any logical SIAs and content-general pragmatic factors that may

[27] Indeed, it is difficult to create an implicit *not* without complexifying the story: 'The villager gave the farmer peanuts' only means he didn't give the farmer corn if you know the villager gave the farmer only one thing. And, while the villager did not live up to his exact agreement, has he acted entirely in bad faith?

exist, causing domain-appropriate inferences to be made even when these violate rules of logical inference.

Indeed, it is exactly this phenomenon of short-circuiting and its explanation – the principle of pre-emptive specificity – that provides the key to understanding how $RT_{General}$ and our views might be mutually consistent. After exploring some of the new design features of evolved specializations illuminated by these experiments, we will explore how relevance theory can be extended to accomplish this.

5.1. Social contract and hazard management algorithms: some new design features

The primary focus of this article has been testing RT_{Wason}'s claim to explain many well-known content effects that otherwise are interpreted to support the hypothesis that there exist social contract and hazard management reasoning specializations. But another positive thread running throughout has been to further explore whether reasoning about social contracts and precaution rules is governed by the same cognitive mechanisms, or by two independent and distinct sets of mechanisms. If progress in mapping the cognitive specializations present in the human mind is to be made, then a careful, progressive mapping of each component is required.

Social contracts and precaution rules share some similarities, e.g. all social contracts are deontic and some precaution rules are; both involve utilities; violations of either can lead to negative outcomes. For this reason, other theorists have failed to sort instances into these two sets, and have accordingly proposed that both sets are processed by the same inferential machinery (although theorists differ in how they characterize that machinery; e.g. Cheng & Holyoak, 1989; Liberman & Klar, 1996; Manktelow & Over, 1995; SCG). In contrast, we have proposed that these two classes of rules are psychologically different: that social contracts and precaution rules are mapped onto different, content-specific representations and evoke different, neurally distinct, content-specialized inference procedures. This second, content-specific view is supported by the discovery of both neural and functional dissociations in reasoning about these two domains (Cosmides & Tooby, 1997; Fiddick, 1998; Stone et al., 1996, 2000). In addition, the content-specific view accurately predicted the outcome of each of the three experiments reported herein.

Indeed, the results of Experiments 1–3 illuminated several design features of the proposed social contract and hazard management algorithms that had not previously been tested. Specifically,

(i) Experiment 1: the concept of 'benefit to X' – a representation of a person's wants or desires – is central to interpreting a situation as involving social exchange and to cheater detection. When mutual desires are expressed in the right kind of context, the linguistic form of a conditional rule is not necessary for the situation to be interpreted as a social contract or to elicit cheater detection. (It is also not sufficient: see Cosmides & Tooby, 1992 on removing benefits from permission rules.)

(ii) Experiments 2 and 3: as predicted by hazard management theory, subjects not only possess an abstract representation of the conditions in which a person is in danger ('engaging in hazardous activity' and 'precaution not taken'), they select the

appropriate cards for discovering who might be in harm's way, even when faced with a hazard management rule that does not express a denial and is *not deontic* in any ordinary sense. To appreciate this result, one must recall that it stands in marked contrast to results involving content-general notions of violation. Subjects do possess an abstract representation of a logical violation – a true antecedent paired with a false consequent – and can recognize such pairings as violating a descriptive conditional rule. Yet they do not select the cards corresponding to these categories when they are asked to *search* for violations. In other words, knowing what counts as a 'violation' is no assurance that there are computational procedures designed to search for them. In this light, the fact that subjects *search* for individuals who may be in danger suggests the operation of a checking subroutine specialized for this task.[28]

(iii) Experiment 3: certain rules can be interpreted as either social contracts or hazard management rules. Using a rule of this kind, we showed that subjects distinguish between social contract and hazard management interpretations, and reason differently (and appropriately) in response to each. Even though exactly the same rule was used in both conditions, totally different patterns of performance were triggered depending on which content-specific concept of violation was cued (cheating versus endangerment). This suggests that each concept plays a central role in representing states of affairs in each domain, and in triggering inferential procedures distinct to each domain. This dovetails with the results involving functional and neurological dissociations (Cosmides & Tooby, 1997; Fiddick, 1998; Stone et al., 1996, 2000), suggesting that despite some superficial similarities, social contracts and precaution rules are not processed by the same reasoning mechanisms.

The results indicating that these two classes of rules are psychologically different also help clarify issues raised by Liberman and Klar (1996). Like SCG, Liberman and Klar argue that the psychological mechanisms activated by the selection task are content-general. In support of this claim, they demonstrated that subjects will continue to answer '*P & not-Q*' on a 'social contract' problem even after the cheating options have been removed (Liberman & Klar, 1996, Experiment 1). But in their revisions of the Gigerenzer and Hug (1992) cholera problem (originally from Cheng & Holyoak, 1985), they unwittingly converted a social contract problem into a hazard management problem, thereby confounding their results.[29] We agree that there are conditions under which people will answer '*P & not-Q*' in response to conditionals not involving social exchange, hazard management rules being one example. But this fails to demonstrate that the inferential machinery responsible

[28] The overactivation of this checking subroutine may play a role in explaining key symptoms of obsessive-compulsive disorder, many of which can be seen as compulsive checking to make sure precautionary measures have been taken: re-checking locks and stoves, repeated hand-washing (an anti-disease measure), and so on (Cosmides & Tooby, 1999).

[29] In the Liberman and Klar (1996) no-cheating-perspective cholera problem, the scenario "cued the subject into the perspective of an Israeli journalist who is interested in risk-taking behavior among young traveling Israelis. He would like to interview youths who violate the rule, which was advised by the Israeli Health Ministry for travelers to the Philippines" (p. 153). This is clearly a precautionary scenario and so would be expected to result in high levels of *P* and *not-Q* selections, according to hazard management theory.

is content-general. The same outcome can be produced by different causes (a headache can be caused by stress or a brain tumor). But if two causes are involved – in this case, two mechanisms – conditions will exist in which the outcomes they produce will differ, as in Experiment 3 above.

5.2. Common ground between ecological rationality and relevance theory: implications for communication and culture

Sperber and Wilson (1995) argue that the human cognitive architecture was sculpted by evolutionary forces into a form whose components tend to operate to maximize relevance (that is, expected cognitive effects) – what they call the *cognitive principle of relevance*. As they say, "human cognition, being an evolved and adapted system, reflects in fine-grained aspects of its design repeated past pressures towards optimisation" (p. 266). Sperber has also eloquently argued that the human mind evolved many domain- or content-specialized cognitive devices whose operations permeate human thought (Sperber, 1994). Most recently, he has advanced a proposal for how selection pressures operating specifically in the context of communication may have led to the evolution of certain logical abilities – a proposal we find very interesting. We too believe that the human cognitive architecture contains some evolved procedures that are identical to (or closely resemble) certain logical procedures, alongside a large set of evolved inference procedures that are ecologically rational but extralogical in structure (Cosmides & Tooby, 1996b). *Modus ponens* would be an example of a logical operation that seems effortless and automatic to all normal subjects.

What is striking is that relevance theory and an ecological rationality perspective, though starting from different problems (communication, reasoning), each converged on conclusions that are in strong agreement with each other on almost every basic point. In particular, both are concerned with explaining how minds that could potentially become sidetracked among an endless immensity of branching inferential chains actually manage so often to settle on the tiny subset of useful ones, effectively pruning unproductive chains before they inflict too much cost (Sperber & Wilson, 1986, 1995; Tooby & Cosmides, 1992). This is a critical engineering issue whether the problem is achieving successful communication, achieving successful social coordination (as in social exchange), achieving the successful acquisition of culture, or indeed the adaptive regulation of behavior generally.

We think this common ground can be enlarged by considering that the principle of pre-emptive specificity (i.e. that specialized inferential machinery will tend to override more general inferential machinery when they are both activated by the same input) might easily be seen to be implied by Sperber and Wilson's two principles of relevance – the cognitive principle of relevance and the communicative principle of relevance (Sperber & Wilson, 1995). Deliverances from a mechanism that specifically evolved to handle a narrow kind of input, and that has received that input, are more likely to be relevant (in either of Sperber and Wilson's two senses) than the simultaneous deliverances from a mechanism that is activated by a far broader range of inputs. Hence, a well-engineered evolved design would often reflect relevance through organized pre-

emption. The same reasons that make at least some logical procedures likely to be reliably developing features of the human cognitive architecture (i.e. that they are widely applicable to many inferential problems) make them likely to be eclipsed by more powerful and internally organized domain-specific machinery, when the inputs fall into the domain governed by a more powerfully focused inference engine. This is why relevance theory can be correct, while RT_{Wason} – whose predictions derive strictly from the deliverances of logical abilities – is false or suspended within the domains of social exchanges and precautions.

This conclusion is good news for reasoning research into content-sensitive capacities, because it means that such capacities will tend to pre-empt other inferential processes within the scope of their proper domains, and hence produce patterned outputs that are robust, replicable, and strongly organized, just as social exchange and hazard management effects appear to be. This means that reasoning experiments can effectively map the design features of content-sensitive and domain-specific inference engines, without results being fragile or easily disrupted by extraneous factors. If pre-emption is a robust phenomenon, then within evolutionarily organized domains, predictions can be clear, precise, and directly tested.[30] One need not be persuaded by Fodor's pessimistic view that what he calls central processes will be forever beyond our ability to study, because of their chaotically interactive complexity (Fodor, 1983).

Not only do we think that the human mind is designed to reflect pre-emption, but we think that the cognitive components of the mind are designed to presume and depend on pre-emption in communication, culture acquisition, and social coordination. Obviously, the human mind uses and is full of contingent, transient, acquired information, and the immense number of free parameters necessary to specify such transient knowledge is one of the great impediments to achieving relevance and avoiding combinatorial explosion (Cosmides & Tooby, 2000). Nevertheless, if the human mind is also permeated by a battery of evolved, content-specialized inference engines, then these are likely to play a leading role in how two or more minds converge on what Sperber and Wilson call the mutually manifest, which they persuasively identify as the indispensable step in successful communication. While we suspect that they would agree with the view we are setting out in this section (and consider it implicitly or explicitly reflected in their writings), we think it is easy to underestimate just how strong the organizing effects of evolved specializations are compared to the deliverances of transient information operated on by logical abilities.

We believe that what stabilizes cultural transmission and even ordinary communication (which are closely intertwined) is that sender and audience can both be certain that the other mind has the same species-typical battery of evolved inference mechanisms. Partners in communication may or may not have the same transient

[30] It is bad news for the study of mental logic, since it means that logical procedures will easily be pre-empted by extralogical procedures, making their presence and properties harder to study. Such a picture would provide an explanation for Rips' otherwise puzzling findings that various logical procedures appear to be off-line much of the time (Rips, 1994).

knowledge, but they can be implicitly counted on to have the same species-typical knowledge. If the principle of pre-emptive specificity is true, this means that speakers, in choosing among alternative forms of expression, should be designed to disproportionately invoke such mechanisms in their audiences to block ambiguity, fix meanings, and to flesh out communication. Reciprocally, interpreters should be designed to expect and to disproportionately fall back on the deliverances of such specializations in attempting to make sense of noisy or confusing situations, using pre-emption to prune or truncate chains of hypotheses about communicative intent that are inconsistent with them. This is why our minimalist Wason task in Experiment 1, in which two parties simply stated what they wanted, was easily fleshed out by subjects into a mutually manifest full scale social exchange complete with dozens of unstated entailments.

A similar argument can be applied to the communicative requirements inherent in games of social coordination. As Nozick (1963) was the first to point out, there is an infinite regress problem in game theory posed by the requirement that players know each other's assumptions, including assumptions about what each player assumes about the other's knowledge about the reciprocal player. Implicit reliance on what is shared species-typically solves this problem, cutting through the regress, allowing social coordination such as social exchange, even with imperfect communication and finite cognitive resources.

It is implicit in Sperber and Wilson's communicative principle of relevance that evolved specializations will provide many of the common reference points and shared conceptual vocabulary that can then be flexibly deployed to deflate ambiguity and combinatorial explosion in the process of communication.

5.2.1. What's mutually manifest? Pre-emption and the origin of background assumptions

How might these considerations apply specifically? It is tempting and plausible to think that the background assumptions that give rise to Wason problem interpretations or relevance judgments are created by culturally acquired knowledge rather than evolved, domain-specific expert systems. But such seemingly self-evident claims can and should be tested. The possibility that evolved, domain-specific expert systems snap subjects into pre-organized interpretive frameworks provides a competing hypothesis for the origin of background assumptions. After all, the fact that many assumptions are derived from cultural experience does not mean that all or even most are.

To illustrate, consider the case of social contract laws, where the weight of evidence supports the hypothesis that the relevant background assumptions are being provided by evolved machinery rather than culturally acquired knowledge. SCG's analysis of the drinking age problem implied that culturally-specific experience with alcohol was the source of the background assumption that there will be beer-drinkers in the absence of the drinking age law. But one needs to be careful. The same assumption can be deduced from the representation of beer drinking as something that is widely desired – something perceived by many as a *benefit*. Moreover, this is the most parsimonious explanation that is consistent with the full range

of data involving social contract laws. After all, high levels of 'P & not-Q' responses are elicited by any law that can be represented as 'If you take the benefit, then you must satisfy the requirement' (and in which a violation involves an illicitly taken benefit) – even by fictious laws set in nonexistent cultures relating exotic elements, such as 'If a man eats cassava root, then he must have a tattoo on his face'. Clearly, American subjects have no prior knowledge of the distribution of imaginary cassava root eating by anyone, tattooed or no. Therefore, culturally-specific knowledge cannot be the source of the background assumptions required for an RT explanation of performance on this social contract law.

Culture learning – if it is to be any kind of a theory at all – must predict, at a minimum, that culturally familiar contents should be treated differently from unfamiliar contents. Yet subjects encountering this odd and culturally unfamiliar cassava–tattoo law for the first time perform just as well as they do on the familiar and culturally overlearned drinking age problem (Cosmides, 1985, 1989).

This effect is extraordinary. It has been replicated many times with different unfamiliar contents, yet its significance remains unappreciated. It is contrary to what virtually all domain-general acquisition theories would predict: on social contract problems, there is no evidence for improved performance even with massively increased exposure. Who would have thought that subjects would do as well on the first exposure to a rule as on the thousandth? What kind of learning curve does this imply?

What allows subjects to deduce that cassava root would be widely eaten in the absence of the law is that (i) it is described in the text as a widely desired benefit, and (secondarily) (ii) intentional agents are making a law to ration access to it. These inferences are made by the joint operation of theory of mind mechanisms (Baron-Cohen, 1995; Leslie, 1994) and social contract algorithms (which pertain to the rationing and regulation of benefits). On this view, all the culture provides is a parameter value: the knowledge of which items and affairs count as benefits. And if the culture does not provide this information, the text of the problem can, leading to the same results. Indeed, by following this logic, cheater detection was elicited from Shiwiar hunter-horticulturalists in the Ecuadorian Amazon, using social contracts created for use with American undergraduates (Sugiyama et al., 2000). (A similar analysis would apply to precaution rules.)

This is another case in which selection task results revealed the operation of evolved SIAs. Without these results, one would mistakenly assume that culture-specific knowledge was providing the relevant background assumptions. As long as RT depends on background assumptions to explain what counts as relevant in a selection task, the selection task has the potential to illuminate those assumptions and, thereby, the inferential machinery that makes them. This is true not only for social exchange and hazard/precaution rules, but across domains. For example, evolved specializations for force dynamics (e.g. Talmy, 1988) provide the background assumptions that allow a relevance theoretic explanation of the Almor and Sloman (1996) matador problem, which employs a descriptive rule.

5.3. The uses of the Wason selection task

SCG recommend that the selection task be abandoned as a tool for studying reasoning. They argue for this as follows: (i) "relevance theory explains the selection task" – that is, performance on all versions of the selection task can be accounted for using logical SIAs and no others; (ii) selection task results do not allow one to choose between alternative theories of the reasoning mechanisms that perform these deductive inferences (e.g. between mental models (Johnson-Laird & Byrne, 1991) and mental logic (Rips, 1994)); (iii) since only logical SIAs are needed to explain performance, the selection task can reveal nothing about the presence of other SIAs – including content-specialized ones – in the human cognitive architecture.

We have shown that, contrary to SCG's claim, logical SIAs are not sufficient to explain performance on many types of selection task. Many non-logical (but adaptively sound) inferences were made by subjects in the selection task experiments presented or discussed herein. Obviously, any inferences that cannot be accounted for by reasoning mechanisms that compute logical equivalences *must be made by some other kind of reasoning mechanism*. Therefore, to the extent that the selection task reveals that people are making non-logical inferences, it does indeed inform one about the presence and structure of extralogical reasoning mechanisms.

5.3.1. A challenge for RT

In SCG's selection tasks, the pruning of background assumptions is accomplished by brute force: the relevant background assumption is explicitly asserted, and the conditional stated in order to deny it. Explicit assertion solves one combinatorial explosion problem: it tells the subject which of a multitude of potential background assumptions are at issue. The context of a disagreement solves a different problem: knowing the speaker's intent in uttering the conditional (i.e. to dispute the assertion). The assertion/denial structure of a dispute is a pragmatic context that has the potential to elicit high levels of '*P & not-Q*' responses almost regardless of a selection task's content: it may indeed be a truly content-general pragmatic device for eliciting falsifying responses (we are agnostic about whether any other conversational device with similar properties exists).

When the conversational device of denial in a dispute is not used, can RT still predict and explain high levels of '*P & not-Q*' responses on descriptive problems? This remains to be seen.

A good test would involve a domain such as intuitive physics/force dynamics. This domain activates an evolved specialization that creates strong, mutually manifest expectations about what will happen when objects interact. This makes it easy to construct a conditional rule whose violation would be surprising given these expectations. Moreover, this specialization (unlike social contract algorithms) presumably lacks a subroutine specialized for detecting violations of the expectations it generates, thus allowing an unconfounded test of RT_{Wason}. At the same time, the pragmatic context would have to be controlled so that it does not imply a disagreement or dispute. High levels of '*P & not-Q*' must be elicited purely by the fact that a

violation would be surprising (i.e. against expectation) and, therefore, productive of cognitive effects.

5.3.2. The benefits of using the selection task

The experiments and analyses presented herein show that the Wason selection task remains a useful and sensitive tool for exploring the computational design of content-specific inferential machinery. Clearly one must control for conversational pragmatics in using this tool, but this is not difficult: we were able to construct several experiments in which the pragmatics of RT_{Wason} led to different predictions from social contract and hazard management theory. Moreover, the need to control for conversational pragmatics is not unique to the selection task; it is required of any verbal task used to probe reasoning, or indeed any experiment that involves instructions or the experimenter communicating with the subject. Indeed, one hopes that all reasoning researchers, regardless of what task they use, will consider SCG's arguments about the effects of conversational pragmatics on reasoning. If anything, SCG have made the selection task more useful: because of their work, we may have a better understanding of how to control for pragmatics in this task than in any other reasoning task.

Moreover, the selection task has special properties that make it an excellent tool for studying how people interpret and reason about conditional rules. (i) It has a clear, simple structure, which makes it particularly easy to control for extraneous factors. (ii) It is simple to substitute different kinds of content into this structure, to see how content affects interpretation and reasoning. There is something to be gained by holding the logical surface structure constant – you get to see what inferences and assumptions the subject imports into the task. (iii) When it comes to the detection of any kind of violation (cheating, conditions of endangerment, logical violation, and others), the selection task has a proven record for uncovering dissociations between knowledge of what counts as a violation, on the one hand, and mechanisms that govern the search for information relevant to the discovery of violations, on the other. This is important, because mechanisms that streamline the search for relevant information should be a design feature of many different adaptive problem-solving systems. (iv) Because the selection task is quick to administer and simple to code, far more experiments are possible than with many other methods, allowing more alternative hypotheses to be evaluated. (v) There is a very large database of selection task experiments against which performance on any new one can be compared. This last advantage is perhaps underappreciated, and is certainly underutilized. The inferential potential of any experiment is higher, the larger the body of results on similar experiments against which it can be compared. There have been so many studies parametrically varying both logical and contentful features of the selection task, that the range of viable reasoning theories consistent with this body of research is highly constrained. This is a powerful benefit, available to anyone willing to master this literature.

Indeed, SCG's own experiments belie the notion that the selection task does little to illuminate human reasoning. Their theory relies on the presence of logical SIAs, and their experiments support the view that there are conditions under which people

will seek evidence that could falsify a conditional rule, even when it is descriptive. This undercuts prior interpretations of selection task results: people's failure to choose the *not-Q* card on descriptive rules had previously been seen as evidence against mental logic theories. Ironically, while SCG themselves doubt the selection task's utility, their own selection task experiments will probably generate a renewed interest in the role of logical form and logical inference in human reasoning.

6. Conclusion

Relevance theory, despite Sperber's own views on the subject (Sperber, 1994), has sometimes been seen as an alternative to social contract theory and other theories positing content-specialized inferential machinery. Moreover, it has in fact been advanced by SCG as a counter-explanation for some of the evidence used to support social contract theory. According to this view, content-general pragmatic factors and logical inferential abilities alone are sufficient to explain performance on the Wason selection task. (If true, such a view would not preclude the existence of social contract algorithms – it would simply cut away at the evidentiary base.)

Contrary to this view, experimental evidence shows that there are many cases in which performance cannot be predicted or explained by SCG's version of relevance theory. Social contract algorithms need to be invoked to explain subject performance on social contracts, and hazard management algorithms for precaution rules. These domain-specific systems supply two necessary elements: (i) an interpretative system consisting of privileged representations and rules of transformation, and (ii) post-interpretive inference procedures, specialized for detecting cheaters or individuals in danger. In the case of social exchange, for example, once a rule is recognized as belonging to the category *social contract*, it is assigned an abstract benefit/requirement representation, and domain-specific rules of implicature are applied (Tables 2 and 3), which specify which translations and transformations are allowable. These license inferences that violate logical constraints, but allow the subject to go beyond the information given in the surface wording of the conditional rule. Post-interpretive processes of cheater detection are then applied to the representation so derived.

Moreover, the data support the principle of pre-emptive specificity: that when more than one reasoning mechanism is activated, the one delivering more content-specific inferences will short-circuit those of more general application. On social contract and precaution rules, subjects consistently detected cheaters and individuals in danger (respectively), even in experiments where this led to different choices than one would predict on the basis of logical rules and content-general pragmatic factors.

It has become commonplace to note that how people reason about a conditional rule depends on how they have interpreted it. And many cognitive scientists have noted that in assigning an interpretation to a rule, people often go beyond the logical form given by the rule's wording. Unfortunately, however, many cognitive scientists

have treated the process of interpretation as a black box, without trying to understand the semantic and pragmatic factors that cause people to interpret rules (and real life situations) in various ways (e.g. Johnson-Laird, 1983). SCG have done us all a great service by trying to pry open that black box and specify some of the interpretive processes that it contains. Surely they are correct in numbering pragmatic factors, such as denial, and logical SIAs among them. But to account for human reasoning performance, the box needs to be stocked with far more. Social contract theory and hazard management theory are accounts of a few of the many additional expert systems that need to be added to this box. They are not just accounts of the post-interpretive reasoning processes that cause the final selection of cards in the Wason task. They are, equally, theories of the processes by which rules that tap into these domains are interpreted. They posit privileged representational systems, and rules of transformation that sanction inferences about what a conditional rule from that domain can imply. They provide a computational account of the common sense by which we spontaneously and intuitively arrive at interpretations in these realms of human life.

While relevance theory is correct in emphasizing the role of interpretive processes in reasoning, its most radical claim – that once content-general pragmatic factors are taken into account, representations of logical form and logical inference procedures are sufficient for understanding performance on the selection task – must be rejected. The evidence from a wide variety of sources – laboratory studies, cross-cultural data, neurological dissociations, and developmental studies – indicates that how people interpret and reason about situations is regulated by a multiplicity of content-specialized spontaneous inferential abilities, such as social contract algorithms and hazard management algorithms, each of which is designed to operate in a distinct domain of human experience. There is no exemption when that experience comes in the form of a Wason selection task.

Acknowledgements

This research was supported by a Natural Sciences and Engineering Research Council of Canada postgraduate scholarship awarded to Laurence Fiddick, by grants to John Tooby from the James S. McDonnell Foundation and the National Science Foundation (#BNS9157-449), and by a Research Across Disciplines Grant from the Office of Research of the University of California, Santa Barbara (Evolution and the Social Mind). We warmly thank Russ Revlin, Valerie Stone, Dan Sperber, an anonymous reviewer, Brad Duchaine and the rest of the UCSB Center for Evolutionary Psychology research group for valuable comments and suggestions on drafts of this paper. We also thank Charles Crawford for his support while the first draft was being written.

Appendix A. About two misreadings of social contract theory

A.1. About costs

Some confusion has persisted in the literature about the role of costs in social contracts, perhaps due to the failure of various commentators (see, e.g. Cheng & Holyoak, 1989) to consult or cite the only publications which actually present social contract theory (Cosmides, 1985; Cosmides & Tooby, 1989). The presence of a cost is not, and never has been, a defining feature of a social contract. The computational theory of social exchange (Cosmides & Tooby, 1989) derives non-arbitrarily from evolutionary analyses (e.g. Axelrod & Hamilton, 1981; Cosmides & Tooby, 1989; Tooby & Cosmides, 1996; Trivers, 1971), not from folk notions. Accordingly, we defined social exchange as cooperation for mutual *benefit* – not the imposition of mutual costs (e.g. Cosmides & Tooby, 1989; Tooby & Cosmides, 1996). For you to offer or agree to a social contract, the situation created must benefit you – this is a necessary condition – but it need not impose a cost on you or anyone else. Agents impose requirements on others in order to create situations that benefit them, and they can be expected to do this *whether or not satisfying that requirement imposes a cost on others*. Thus, obviously it is reciprocal conditional benefits and not the presence of a cost that defines a social contract (*pace* Cheng & Holyoak, 1989). According to the computational theory, one may well incur a cost in satisfying a requirement, but this is neither a necessary nor a sufficient condition for a conditional to express a social contract (see also Tooby & Cosmides, 1989, 1996).

Of course, at an earlier stage in processing, in deciding whether to accept an offer, one must always evaluate whether the benefits outweigh the costs of fulfilling the requirement. This is done by subtracting the cost of fulfilling the requirement from the expected benefit, to see if the result is sufficiently positive. This requires that the same element be represented in two different ways – as a requirement, and as a price. Because of this processing sequence, we sometimes refer to the requirement as the cost term, or the cost/requirement term, with the unfortunate side-effect that those who use their folk notions as opposed to consulting the computational theory have been misled into thinking it is a part of the theory that the requirement term necessarily inflicts a cost. The requirement may as a byproduct often be costly, but it is not important (much less necessary) to the exchange relationship that it be so. It is also true that, in a highly ambiguous situation, the costliness of a requirement may be one cue that a social contract is being offered (this has been the basis for certain experiments, e.g. Cosmides, 1989; Platt & Griggs, 1993). Nevertheless, this cue should be and is ineffective if it is difficult to interpret anything in the rule as involving a benefit, because benefits are the essential elements of the exchange.

A.2. About laws

Consulting their intuitions rather than social contract theory, Cheng and Holyoak (1989) make a sharp distinction between social contracts that involve the exchange of goods and ones that restrict access to a good on the basis of satisfying a require-

ment. There is no theoretical justification for this on evolutionary grounds (and they provide none). The domain of social contracts encompassed more than just the exchange of goods during hominid evolutionary history, and indeed, to judge by primate studies, objects may have been included in exchanges at a relatively late stage (Cosmides, 1985; Cosmides & Tooby, 1989; Tooby & Cosmides, 1996). For example, two hunter-gatherer bands, the Gana and the !Kung, may have a long-standing agreement about water hole privileges as part of a larger system of reciprocity: from the !Kung band's point of view, there is a social contract 'If you use our water hole, you must be a member of the Gana band' (Shostak, 1981). Being a member of the Gana band is not a good that is exchanged, nor is it costly in the everyday sense of the word. But this situation fits squarely into the definition of a social contract as specified by our computational theory of social exchange (Cosmides, 1985; Cosmides & Tooby, 1989). It is one part of an agreement to cooperate for mutual benefit. Similarly, a social group may restrict access to a benefit (such as alcohol) to people of a specific age (a proxy for responsibility), because this creates a situation that benefits them (they are safer when drinking is restricted to older people); again, this fits the structure of a social contract by our definition (see also footnote 24).

Appendix B. Instructions and RT_{Wason}

Instruction effects provide little explanatory force for RT_{Wason}, because invoking them to explain results in one case will invalidate arguments used to explain contrary results in other, similar cases.

For example, to use instruction effects to explain the results of Experiment 2, an RT_{Wason} argument would have to go something like this: in the *Precaution* version, subjects were asked *whether people are endangering themselves*. Although 'endangerment' is not itself a logical concept, the text suggests that the reason some of your fellow tribesmen might be in danger is because they 'might not know about the jackets'. A person who does not know about the jackets cannot be expected to be wearing them under the proper conditions. This is equivalent to suggesting that some people might – however unwittingly – be 'violating' the conditions stated for use of the jackets. This could engage a falsification strategy, thereby highlighting the disconfirming conjunction of features, *P-and-(not-Q)*. In contrast, when the *Standard* version asked subjects to determine *whether the rule is true*, this could engage a verification strategy, thereby highlighting the confirming conjunction of features, *P-and-Q*. If so, then subjects would be more likely to choose the *not-Q* card in the *Precaution* version than in the *Standard* version.

The first problem with this explanation is that, according to RT_{Wason}, instructions that highlight violations can have no effect on performance unless the subject also makes inference (c). Otherwise, RT_{Wason} has no explanation for the robust finding that violation instructions fail to improve performance for ordinary descriptive rules. But the scenario in Experiment 2 should not have triggered inference (c). In making the latter prediction, we followed the reasoning laid out by SCG: it should be

difficult to interpret a descriptive conditional as denying prior expectations if there are none. This puts RT_{Wason} into an explanatory double bind. To explain the results, RT_{Wason} would either have to abandon its elegant explanation for why violation instructions have no effect on descriptive rules not involving denial (leaving that phenomenon unexplained) or it would have to abandon its claims about the conditions under which inference (c) is made (relinquishing some of its most precise claims about the importance of conversational pragmatics).

Moreover, even if there were a way to tweak or reconform RT_{Wason} such that it explained how inference (c) could be produced by these problem contents, this would nevertheless simply substitute one explanatory difficulty for another. To see why, consider stretching RT_{Wason} to argue one of the following:

(i) Although the asker had no idea what the orange jackets were for, because they are new his default assumption is that people don't wear them during *any* activity (which would include hunting). Although the asker did not know hunting was at issue, the person making the reply did, and he also knew that the asker knows that people have always hunted without orange jackets in the past (i.e. that instances of *P-and-(not-Q)* normally occur). Hence, his reply was meant to deny this (deeply implicit and unstated) background assumption. Or,

(ii) The descriptive rule might remind subjects of precautionary deontic rules. Because these are interpreted as forbidding cases of *P-and-(not-Q)*, subjects invalidly import this interpretation into a problem that does not, on its own, warrant it. (This may stretch the theories too much, in that it renders both RT_{Wason} and every deontic theory of Wason performance virtually unfalsifiable.)

In either case, inference (c) would be made. But that is also the problem. The accounts given in (i) and (ii) apply equally to the *Standard* and the *Precaution* version. If inference (c) is drawn for both, then subjects should choose the *not-Q* card for both – which they did not do. After all, in SCG's own demonstrations, subjects selected the *not-Q* card whenever inference (c) was warranted by the pragmatic context. This occurred *without any violation instructions*; indeed, it occurred even when the instructions asked subjects to determine whether the stated rule was *true*: an instruction that highlights verification over falsification. So, the difference in instructions between the two conditions does not allow RT_{Wason} to explain the difference in subject response.

References

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

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

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

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

Baillargeon, R. (1986). Representing the existence and the location of hidden objects: object permanence in 6- and 8-month old infants. *Cognition*, *23*, 21–41.

Barkow, J., Cosmides, L., & Tooby, J. (1992). The adapted mind: evolutionary psychology and the generation of culture. New York: Oxford University Press.

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

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.

Brown, A. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science*, *14*, 107–133.

Caramazza, A., & Shelton, J. (1998). Domain-specific knowledge systems in the brain: the animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*, 1–34.

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

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

Cheng, P., Holyoak, K., Nisbett, R., & Oliver, L. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, *18*, 293–328.

Chrostowski, J., & Griggs, R. (1985). The effects of problem content, instructions, and verbalization procedure on Wason's selection task. *Current Psychological Research and Reviews*, *4*, 99–107.

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 No. 86-02206)

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

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

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

Cosmides, L., & Tooby, J. (1996a). Are humans good intuitive statisticians after all?: rethinking some conclusions of the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.

Cosmides, L., & Tooby, J. (1996b). A logical design for the mind? *Contemporary Psychology*, *41* (5), 448–450.

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

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

Cosmides, L., & Tooby, J. (2000). Consider the source: the evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentations: a multidisciplinary perspective*. Vancouver studies in cognitive science. New York: Oxford University Press.

Cosmides, L., Tooby, J., Montaldi, A., & Thrall, N. (1999, June 2–6). *Character counts: cheater detection is relaxed for honest individuals*. Paper presented at the Human Behavior and Evolution Society, Salt Lake City, UT.

Cummins, D. (1996). Dominance hierarchies and the evolution of human reasoning. *Minds and Machines*, *6*, 463–480.

Cummins, D. (1998). Social norms and other minds: the evolutionary roots of higher cognition. In D. D. Cummins, & C. A. Allen (Eds.), *The evolution of mind*. New York: Oxford University Press.

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

Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Evans, J. St. B. T. (1982). *The psychology of deductive reasoning*. London: Routledge and Kegan Paul.

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

Fiddick, L., Cosmides, L., & Tooby, J. (1995). *Are there really separate reasoning mechanisms for social contracts and precautions?* Paper presented at the Seventh Annual Meeting of the Human Behavior and Evolution Society, University of California, Santa Barbara, CA.

Fiddick, L., Cosmides, L., & Tooby, J. (2000). Does the mind distinguish between social contracts and precautions? Dissociating cognitive adaptations through inference priming. Manuscript submitted for publication.

Fiske, A. (1991). *Structures of social life: the four elementary forms of human relations*. New York: Free Press.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
Gallistel, C., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74.
Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.
Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, *102*, 684–704.
Gigerenzer, G., & Hug, K. (1992). Domain specific reasoning: social contracts, cheating, and perspective change. *Cognition*, *43*, 127–171.
Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.
Gigerenzer, G., & Todd, P. of the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
Griggs, R. (1984). Memory cueing and instructional effects on Wason's selection task. *Current Psychological Research and Reviews*, *3*, 3–10.
Griggs, R. (1989). To "see" or not to "see": that is the selection task. *Quarterly Journal of Experimental Psychology*, *41A*, 517–529.
Griggs, R., & Cox, J. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*, 407–420.
Gutheil, G., Vera, A., & Keil, F. (1998). Do houseflies think? Patterns of induction and biological beliefs in development. *Cognition*, *66*, 33–49.
Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition*, *50*, 171–188.
Hirschfeld, L. & Gelman, S. (1994). *Mapping the mind: domain specificity in cognition and culture*. New York: Cambridge University Press.
Hoffman, E., McCabe, K., & Smith, V. (1998). Behavioral foundations of reciprocity: experimental economics and evolutionary psychology. *Economic Inquiry*, 36, 335–352.
Jackson, S., & Griggs, R. (1990). The elusive pragmatic reasoning schema effect. *Quarterly Journal of Experimental Psychology*, *42A*, 353–373.
Johnson-Laird, P. (1983). *Mental models: towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
Johnson-Laird, P., & Byrne, R. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
Keil, F. (1994). The birth and nurturance of concepts by domain: the origins of concepts of living things. In L. Hirschfeld, & S. Gelman (Eds.), *Mapping the mind: domain specificity in cognition and culture*. New York: Cambridge University Press.
Kirby, K. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition*, *51*, 1–28.
Kroger, J., Cheng, P., & Holyoak, K. (1993). Evoking the permission schema: the impact of explicit negation and a violation-checking context. *Quarterly Journal of Experimental Psychology*, *46A*, 615–635.
Kurzban, R., Rutherford, M., Cosmides, L., & Tooby, J. (1997). *Cooperation and punishment in groups: economic trade-offs*. Paper presented at the Ninth Annual Meeting of the Human Behavior and Evolution Society, University of Arizona, Tucson, AZ.
Leslie, A. (1987). Pretense and representation: the origins of "theory of mind". *Psychological Review*, *94*, 412–426.
Leslie, A. (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*. New York: Cambridge University Press.
Leslie, A., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*, 265–288.
Liberman, N., & Klar, Y. (1996). Hypothesis testing in Wason's selection task: social exchange cheating detection or task understanding. *Cognition*, *58*, 127–156.
Manktelow, K., & Evans, J. St. B. T. (1979). Facilitation of reasoning by realism: effect or non-effect? *British Journal of Psychology*, *70*, 477–488.
Manktelow, K., & Over, D. (1987). Reasoning and rationality. *Mind & Language*, *2*, 1990–2190.
Manktelow, K., & Over, D. (1988). *Sentences, stories, scenarios, and the selection task*. Paper presented at the First International Conference on Thinking, Plymouth, UK.

Manktelow, K., & Over, D. (1990). Deontic thought and the selection task. In K. Gilhooly, M. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thought: reflections of the psychology of thinking* (pp. 153–164). London: Wiley.

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

Manktelow, K., & Over, D. (1992). Utility and deontic reasoning: some comments on Johnson-Laird & Byrne. *Cognition*, *43*, 183–186.

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

Nozick, R. (1963). *The normative theory of individual choice*. New York: Garland Press reprinted 1990.

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

Pinker, S. (1997). *How the mind works*. New York: Norton.

Platt, R., & Griggs, R. (1993). Facilitation in the abstract selection task: the effects of attentional and instructional factors. *Quarterly Journal of Experimental Psychology*, *46A*, 591–613.

Pollard, P. (1990). Natural selection for the selection task: limits to social exchange theory. *Cognition*, *36*, 195–204.

Reich, S., & Ruth, P. (1982). Wason's selection task: verification, falsification and matching. *British Journal of Psychology*, *73*, 395–405.

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

Rutherford, M., Cosmides, L., & Tooby, J. (1996, June). *Adaptive sex differences in reasoning about self defense*. Paper presented at the Eighth Annual Meeting of the Human Behavior and Evolution Society, Evanston, IL.

Rutherford, M., Cosmides, L., & Tooby, J. (2000) Precaution evaluation and the need for well-formed inputs, in preparation.

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

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

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. Hirschfeld & S. Gelman, *Mapping the mind: domain specificity in cognition and culture* (pp. 39–67). New York: Cambridge University Press.

Sperber, D. (1997, February). *Culture and the epidemiology of meta-representations*. Paper presented at the Tenth Annual Vancouver Cognitive Science Conference, Simon Fraser University, Vancouver, Canada.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: a multidisciplinary perspective*. *Vancouver studies in cognitive science*. New York: Oxford University Press.

Sperber, D., & Wilson, D. (1986). *Relevance: communication and cognition* (1st ed.). Oxford: Blackwell.

Sperber, D., & Wilson, D. (1995). *Relevance: communication and cognition* (2nd ed.). Oxford: Blackwell.

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

Springer, K. (1992). Children's awareness of the implications of biological kinship. *Child Development*, *63*, 950–959.

Stone, V., Cosmides, L., & Tooby, J. (1996). *Selective impairment of cheater detection: neurological evidence for adaptive specialization*. Paper presented at the Eighth Annual Meeting of the Human Behavior and Evolution Society, Northwestern University, IL.

Stone, V., Cosmides, L., Tooby, J., Knight, R., & Kroll, N. (2000). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage, under review.

Sugiyama, L. (1996). *In search of the adapted mind: a study of human cognitive adaptations among the Shiwiar of Ecuador and the Yora of Peru*. Doctoral Dissertation, Department of Anthropology, University of California, Santa Barbara, CA.

Sugiyama, L., Tooby, J., & Cosmides, L. (2000) Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia, under review.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, *12*, 49–100.

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

Tooby, J., & Cosmides, L. (2000). Ecological rationality in a multimodular mind. *Evolutionary psychology: foundational papers*. Cambridge, MA: MIT Press in press.

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

Valentine, E. (1985). The effect of instructions on performance in the Wason selection task. *Current Psychological Research and Reviews*, *4*, 214–223.

Wason, P. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, *4*, 7–11.

Wason, P. (1966). Reasoning. In B. M. Foss, *New horizons in psychology*. Harmondsworth: Penguin.

Wason, P. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*. *20*, 273–281.

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

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

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*, 749–750.

Wynn, K. (1995). Origins of numerical knowledge. *Mathematical Cognition*, *1*, 35–60.

Yachanin, S. A. (1986). Facilitation in Wason's selection task: content and instructions. *Current Psychological Research and Reviews*, *5*, 20–29.