Cognition, 43 (1992) 127-171

Domain-specific reasoning: Social contracts, cheating, and perspective change*

Gerd Gigerenzer and Klaus Hug

Institute of Psychology, University of Salzburg, Salzburg, Austria

Received June 11, 1990; final revision accepted November 14, 1991

Abstract

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

What counts as human rationality: reasoning processes that embody content-independent formal theories, such as propositional logic, or reasoning processes that are well designed for solving important adaptive problems? Most theories of human reasoning have been based on content-independent formal rationality, whereas adaptive reasoning, ecological or evolutionary, has been little explored. We elaborate and test an evolutionary approach, Cosmides' (1989) social contract theory, using the Wason selection task. In the first part, we disentangle the theoretical concept of a "social contract" from that of a "cheater-detection algorithm". We demonstrate that the fact that a rule is perceived as a social contract – or a conditional permission or obligation, as Cheng and Holyoak (1985) proposed – is not sufficient to elicit Cosmides' striking results, which we replicated. The crucial issue is not semantic (the meaning of the rule), but pragmatic: whether a person is cued into the perspective of a party who can be cheated. In the second part, we distinguish between social contracts with bilateral and unilateral cheating options. Perspective change in contracts with bilateral cheating options turns P & not-Q *responses into* not-P & Q *responses. The results strongly support social*

Correspondence to: Gerd Gigerenzer, Institut für Psychologie, Universität Salzburg, Hellbrunnerstraße 34, 5020 Salzburg, Austria.

*This article was written while Gerd Gigerenzer was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, CA. We want to thank Leda Cosmides, Lorraine Daston, Juan Delius, Vittorio Girotto, Wolfgang Hell, Ralph Hertwig, Andreas Krebs, David Over, John Tooby, Paolo Legrenzi, Ken Manktelow, Dan Sperber, and an anonymous reviewer, for helpful comments. We are grateful for the financial support provided by the Spencer Foundation, the Deutsche Forschungsgemeinschaft (DFG 170/2-1) and the Centre for Interdisciplinary Research, Bielefeld.

0010-0277/92/$14.00 © 1992 – Elsevier Science Publishers B.V. All rights reserved.

contract theory, contradict availability theory, and cannot be accounted for by pragmatic reasoning schema theory, which lacks the pragmatic concepts of perspectives and cheating detection.

Introduction

What does it mean to be rational? Leibniz' vision was to reduce human reasoning to a calculus, the Universal Characteristic, which would settle all arguments in a rational way by coercing assent, once everyone accepted the rules. If a dispute arose, the contending parties could settle it by sitting down and calculating. Similarly, Boole saw formal logics and human reasoning as two sides of the same coin; indeed, when he wrote on the foundations of logic, algebra, and probability, he called his treatise *An Investigation of the Laws of Thought* (Boole, 1854/1958). In psychology, research on the human intellect has borne the imprint of the Enlightenment conviction that reasoning could be reduced to a general calculus (on the rise and fall of this idea see Daston, 1988). For instance, when Inhelder and Piaget (1958, p. 305) asserted "Reasoning is nothing more than the propositional calculus itself", they echoed Laplace's statement one and a half centuries earlier: probability theory is "nothing more at bottom than good sense reduced to a calculus" (Laplace, 1814/1951, p. 196).

The century-old conviction that humans reason according to some content-independent logic was shattered by a number of factors, among them research on the selection task introduced by Peter Wason (1966). In the selection task, the subject is asked to search for information that can violate or falsify a conditional rule. The main result of this research is that reasoning is guided by the *content* of the task, rather than by its formal structure. Although it was realized that "content is crucial" (Wason & Johnson-Laird, 1972, p. 245) for understanding how humans actually *do* reason, content-independent logic was retained in the 1970s and 1980s as the yardstick for how subjects *should* reason. Subjects' judgments were examined for their *deviations* from that logic. Deviations were often called "fallacies" or "biases", and attributed to deeper-level deficits in information processing, such as "confirmation bias" and "matching bias". Content was not of theoretical interest in and of itself, but was seen as a factor that could "facilitate" logical reasoning, or hinder it. In this view, the explanandum was the "content effect", that is, why certain contents "facilitated" logical reasoning whereas others did not.[1]

[1] The term "content effect" refers to a percentage of *P & not-Q* responses in a "thematic" rule that is substantially larger than the corresponding percentage in an "abstract", alphanumerical rule such as "If there is a *D* on one side of any card, then there is a *3* on its other side." The percentage of *P & not-Q* responses in "abstract" rules is typically below 20% (Cosmides, 1989; Evans, 1982). Both thematic and abstract rules are conditionals of the form "if *P* then *Q*".

What would a theoretical framework look like that starts with content as a primary concept, rather than as a modifier of logical reasoning? For the selection task, two such proposals exist: social contract theory (Cosmides, 1985, 1989; Cosmides & Tooby, 1989) and pragmatic reasoning schema theory (Cheng & Holyoak, 1985, 1989; Cheng, Holyoak, Nisbett, & Oliver, 1986). Both theories hold that reasoning processes are *domain specific* as opposed to domain general. Both theories model aspects of *deontic* reasoning, reasoning about "must" and "may", obligation and entitlement. They contribute to two current and related debates: a normative one and a descriptive one. First, what counts as human rationality: reasoning processes that embody content-independent formal rules, such as propositional logic, or reasoning processes that are well designed for solving important adaptive problems, such as social contracts or social regulations? Second, how specific are the laws of reasoning? Do they consist of rules that are general purpose and can be applied to any problem, or rules that are domain specific and designed for a limited class of problems?

In this article, we elaborate and test social contract theory. This theory postulates (1) that we should think of reasoning as rational insofar as it is well designed for solving important adaptive problems, and (2) that there exist domain-specific cognitive processes for reasoning about social contracts. In Part I we disentangle the theoretical concept of a "cheater-detection algorithm" from that of a "social contract rule" and provide a novel experimental test of social contract theory. In Part II we introduce the concept of "perspective", the concepts of "unilateral" and "bilateral cheating options", and the use of perspective change to test the explanatory concept of a cheater-detection algorithm.

Throughout, we test the predictions of social contract theory against those of pragmatic reasoning schema theory and availability theory.

Social contract theory

Social contract (SC) theory posits a modular and evolutionary view of human reasoning. "Modular" means that the theory explains performance in one specific content domain: social contracts. Note that earlier explanations such as "confirmation bias" were not domain specific; they were understood to be a general tendency of the mind, which had no specific link to particular contents – just as standard propositional logic and probability theory are not seen as theories about specific contents.[2] What is a social contract? In Cosmides' words, "a social contract relates *perceived benefits* to *perceived costs*, expressing an exchange in

[2] The separation of mathematical probability and statistics from a specific content (such as rationality) is, however, historically a recent development. One could date it as late as 1933, when Kolmogoroff published his axiomatization of probability which finally freed probability from its previous contents (Gigerenzer et al., 1989).

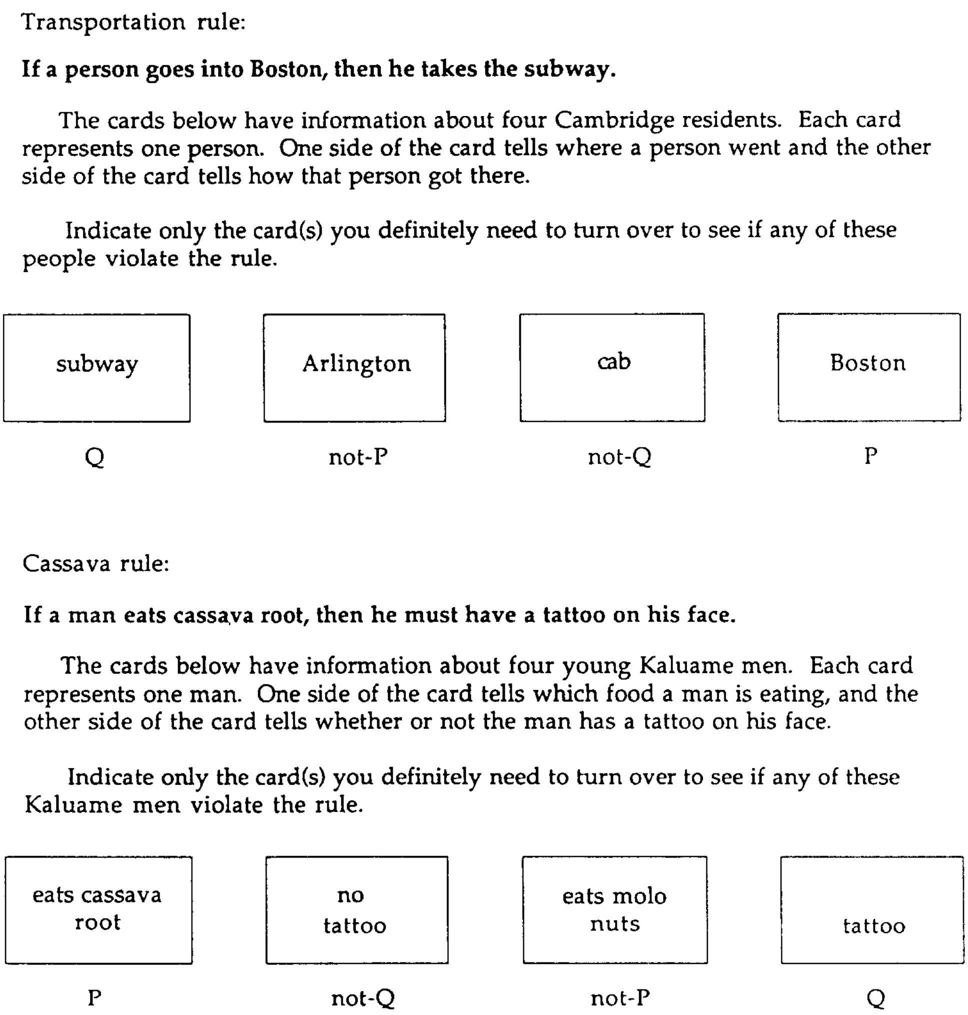
which an individual is required to pay a cost (or meet a requirement) to an individual (or group) in order to be eligible to receive a benefit from that individual (or group). Cheating is the failure to pay a cost to which one has obligated oneself by accepting a benefit, and without which the other person would not have agreed to provide the benefit" (1989, p. 197).

The evolutionary part of social contract theory is, in brief: our species spent more than 99% of its history as Pleistocene hunter-gatherers. For hunter-gatherers, social contracts, that is, cooperation between two or more people for mutual benefit, were necessary for survival. But cooperation (reciprocal altruism) cannot evolve in the first place unless one can detect cheaters (Trivers, 1971). Consequently, a set of reasoning procedures that allow one to detect cheaters efficiently – a cheater-detection algorithm – would have been selected for. Such a "Darwinian algorithm" would draw attention to any person who has accepted the benefit (did he pay the cost?) and to any person who has not paid the cost (did he accept the benefit?). Because these reasoning procedures, which were adaptations to the hunter-gatherer mode of life, are still with us, they should affect present-day reasoning performance. Therefore, argues Cosmides, we should find traces of Darwinian algorithms even in reasoning about textbook problems such as the selection task.

We want to add here that notions similar to a cheater-detection algorithm could also be derived from points of view other than an evolutionary one, such as from the work on children's understanding of deception as a function of their ability for perspective change (e.g., Wimmer & Perner, 1983).

Let us now look at two selection tasks used by Cosmides (Figure 1). The rule in the first selection task read: "If a person goes into Boston, then he takes the subway." We refer to this rule as the "transportation rule". The subjects (Harvard students) who were given this rule were told that their job was to study the demographics of transportation. The four cards had information about four Cambridge, MA, residents. One side of the card told where a person went, the other side how the person got there. The subjects' task was to indicate those card(s), and *only* those cards, one definitely needs to turn over in order to see if any of these people violate the rule. The transportation rule has the formal structure "if *P* then *Q*", and the possible selections are marked in Figure 1 by "*P*", "*not-P*", "*Q*", "*not-Q*". (The order of cards was, as in our experiments, random.)

In propositional logic, a conditional "if *P* then *Q*" is false only if *P* is true and *Q* is false. In the three other possible combinations, the conditional is always true. Hence, in order to find out whether the conditional is logically false, one should turn over the *P* card (to check whether *not-Q* is on the other side) and the *not-Q* card (to check whether *P* is on the other side). Nevertheless, in many studies on the transportation rule only about 30–40% of subjects made the selection *P* & *not-Q* (for an overview see Cosmides, 1989, p. 200). For instance, many subjects chose *P* & *Q*.



Transportation rule:

If a person goes into Boston, then he takes the subway.

The cards below have information about four Cambridge residents. Each card represents one person. One side of the card tells where a person went and the other side of the card tells how that person got there.

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

subway	Arlington	cab	Boston
Q	not-P	not-Q	P

Cassava rule:

If a man eats cassava root, then he must have a tattoo on his face.

The cards below have information about four young Kaluame men. Each card represents one man. One side of the card tells which food a man is eating, and the other side of the card tells whether or not the man has a tattoo on his face.

Indicate only the card(s) you definitely need to turn over to see if any of these Kaluame men violate the rule.

eats cassava root	no tattoo	eats molo nuts	tattoo
P	not-Q	not-P	Q

Figure 1. *The selection task: Transportation rule and cassava rule. (The "P's" and "Q's" do not appear on versions given to subjects.)*

The transportation rule is not a social contract. There are no two partners who have engaged in a contract, nor is *P* a benefit for one partner and a cost for the other, nor does this hold for *Q*. Therefore, SC theory is mute on this problem.

Cosmides then devised a conditional rule to which SC theory did apply: "If a man eats cassava root, then he must have a tattoo on his face" (Figure 1). The cassava rule was explained in a context story as a social contract in a tribe called the Kaluame. Cassava root is a scarce and powerful aphrodisiac. Among the Kaluame, only married men have tattoos on their faces. The elders have established the cassava rule because they strongly disapprove of sexual relations between unmarried people. Many unmarried men, however, are tempted to cheat. Cosmides cued the subjects into the perspective of a guard whose task is to catch persons breaking the law – that is, Kaluame men who violate the cassava rule. Each card had information about one Kaluame man. Again, the subjects' task was to indicate which cards have to be turned over to see if any of the four men violated the rule.

The context story identified the cassava rule as having the structure of a social contract: "If you take the benefit, then you pay the cost (or meet the requirement)". Eating cassava root is a benefit compared with the alternative, molo nuts, and having a tattoo (being married) may be better described as a requirement than a cost. Because the rule is a social contract rule, SC theory makes a prediction: a cheater-detection algorithm is activated. Since cheating means taking the benefit P and not meeting the requirement Q, subjects should select P *& not-Q*. In fact, about 75% of Cosmides' subjects selected P *& not-Q* in this social contract problem.

Although SC theory is particular to the domain of social contracts, the evolutionary perspective presented by Cosmides (1985, 1989) and Cosmides & Tooby (1989) predicts that domain-specific reasoning procedures should exist for other evolutionary important domains (such as threats and warnings) as well.

Pragmatic reasoning schema theory

Cheng and Holyoak (1985) proposed that the content of a conditional rule cues one of several "pragmatic reasoning schemas (PRS)". A "permission schema", for instance, is cued if the rule relates an "action" to a "precondition" in one of four ways. A permission schema specifies a production rule for each of the four possible antecedents (cards) (p. 397):

Rule 1: If the action is to be taken, then the precondition must be satisfied.
Rule 2: If the action is not to be taken, then the precondition need not be satisfied.
Rule 3: If the precondition is satisfied, then the action may be taken.
Rule 4: If the precondition is not satisfied, then the action must not be taken.

Since "taking a benefit" and "paying costs" are all "actions", it seems that all social contracts are permission rules (or obligation rules, see below), but not vice versa. The cassava rule is both a social contract and a permission rule.[3] PRS theory assumes that if a rule has the action-precondition structure of one of the four rules above, then the entire set of four rules composing the schema becomes available. The cassava rule is in the form of rule 1; therefore rule 4 becomes

[3] Cheng and Holyoak's (1985) use of the term "permission rule" for conditionals of the type of Rule 1 has been criticized by Manktelow and Over (1991), because there is a "must" and not a "may" in the consequent of the conditional. Nevertheless, in order to avoid confusion, we will use the terms "permission rule" and "obligation rule" in this article as defined by Cheng and Holyoak.

available. Rule 4 specifies where to look for the potential violation, namely, "*not-Q* must not occur with *P*", or "*P & not-Q*". PRS theory and SC theory give the same prediction for the cassava rule, and that prediction coincides with propositional logic. But the two theories can sometimes give predictions that diverge from one another, and from propositional logic, as we shall show.

The transportation rule, in contrast, does not prescribe a regulation between an action and a precondition. It does not contain the deontic concepts of obligation and entitlement, expressed in English by the modals "must" and "may". Therefore, it is not a permission rule, and the production rules of the permission schema cannot be applied.

SC theory and PRS theory share a domain-specific approach to reasoning. However, they differ in their definitions of what the domain and the reasoning processes are. The debate between Cheng and Holyoak (1989) and Cosmides (1989) suggests that both sides see the essential difference between the two theories in the *semantic interpretation of the rule*: whether a rule is a social contract or a permission (or obligation) rule. Social contracts are a subset of permission (or obligation) rules. The precise difference is blurred, because perceived "benefits", "costs", "actions", and "preconditions" are fuzzy concepts with no clear-cut boundaries. This does not mean that no clear prototypes exist. But the fuzziness has generated a heated debate. In contrast, we argue that the following two notions help to reveal the major differences.[4] (1) PRS theory does not use the notions of *cheating option* and *cheater-detection algorithm*, which are central for SC theory. (2) The notion of cheating implies that (at least) two *parties* with two different *perspectives* exist (Cosmides & Tooby, 1989, p. 85). Being cheated is relative to the perspective of one party, whereas PRS theory has no provision for perspectives in the production rules. We will exploit this difference below.

Availability theories

Most explanations of the effect of content on reasoning in the selection task do not refer to domain-specific reasoning processes. Rather, these attribute the effect of content (any content) to the *amount of experience* a subject has had with this content. These explanations link associationism with Tversky and Kahneman's (1973) "availability heuristic". Note that "availability" explanations are content

[4] These are the differences essential to the tests in Parts I and II of this paper. SC theory makes other claims about inferences (e.g., Cosmides & Tooby, 1989, pp. 79–83), and PRS theory also deals with performance in domains other than social contracts. There are also various kinds of conditional permissions and obligations that have been recently studied by Girotto, Blaye, and Farioli (1989), Girotto, Gilly, Blaye, and Light (1989) and Manktelow & Over (1990, in press), as well as other deontic rules such as conditional precautions (Manktelow & Over, 1990) and promises (Light et al., 1990).

independent; the assumption is that they can be applied to any problem of any content. The role of content in availability explanations is of a different kind: the assumption is that subjects are influenced by their familiarity with the content of a rule.

There are various formulations, but common to all availability theories are the conditions that (1) the subject's past experience has created associations between the propositions in a conditional rule, and (2) the more exposures a subject has had to, say, *P* and *Q*, the stronger that association will be and the more easily *P* and *Q* will come to mind – be "available" as a response (e.g., Johnson-Laird, 1982; Manktelow & Evans, 1979; Pollard, 1982; Wason, 1983). For instance, according to the "memory cueing hypothesis" of Griggs and Cox (1982), "performance on the selection task is significantly facilitated when the presentation of the task allows the subject to recall past experience with the content of the problem, the relationship expressed, and a counter-example to the rule governing the relationship" (p. 417). In one version of the availability theory it is required only that *one* falsifying instance can be recalled from long-term memory in order to produce a *P & not-Q* response (Griggs & Cox, 1982, 1983). In contrast, in the "differential availability hypothesis", *P & not-Q* instances must be more available than, say, *P & Q* instances, to produce this response (Pollard, 1982). In their review of the "thematic-materials" effect ("content effect", i.e., *P & not-Q* responses are more frequent if *P* and *Q* are thematic rather than abstract), Griggs and Cox (1982) concluded with a strong statement about the explanatory power of long-term memory cues: "In summary, the evidence for the thematic-materials effect that cannot be directly attributed to memory-cueing is weak and inconsistent. When the data indicate a strong effect, it can almost invariably be attributed to memory cueing and not facilitation of logical reasoning" (p. 419).

Let us now apply this to the transportation and cassava rules. What does availability predict? First, since availability theory is not domain specific, it can be applied to any rule and therefore to both. Second, cassava root is not familiar to North American and European subjects, nor is the connection between cassava root and tattoo, nor can memory provide counter-examples from experience. Availability theory thus predicts a low percentage of *P & not-Q* responses for the cassava rule. This contrasts sharply with the prediction from both SC and PRS theory. In the transportation rule, however, the propositions, the relation, and even counter-examples are familiar (although familiarity may vary with the subjects' geographical location). In any case, availability theory should predict that there will be far more *P & not-Q* responses in the familiar transportation problem than in the unfamiliar cassava problem. Cosmides, however, found that the reverse holds.

In what follows we elaborate and reformulate SC theory, making conceptual distinctions explicit that were only implicit in Cosmides' (1989) work, and testing these. We first describe the experimental design and procedure we used for all the tests reported later in the paper.

Experimental design

We used a within-subjects design. Each subject received 12 selection tasks. We will explain each of these problems when we present and discuss the results. Here we describe only the general structure of the design.

Subjects

Ninety-three students from the University of Konstanz participated in the experiment. Students were paid volunteers, recruited by advertisement from a broad spectrum of disciplines, including biology, law, linguistics, management, mathematics, and psychology. There were 58 female and 35 male students with a mean age of 22.6 years (standard deviation: 2.9 years, range 19–36 years). None of the subjects had any prior experience with the selection task.

Materials and procedure

Each selection task consisted of a rule, an instruction, and a context story. Rules and instructions in all selection tasks were as shown in Figure 1; context stories were either taken from Cosmides (1989) or constructed by us. There were 12 rules with two versions each, resulting in 24 selection tasks. Two series of 12 selection tasks were constructed, each of which contained one and only one version of each rule (Table 1). Subjects were randomly assigned to one of the two series. Thus no subject got both versions of a rule. Subjects also received six additional selection tasks that did not involve social contracts, which are not analyzed here.

There were two sets of rules, corresponding to the two parts of this paper. Each set had two theoretical goals (see Table 1). In the first set (rules 1–6), a rule had either a "cheating" context story or a "no-cheating" context story. In the second set (rules 7–12), a rule had a context story that cued the subjects either into the perspective of party A or into that of party B (A and B being the parties engaged in a social contract). All context stories were of about the same length as those used by Cosmides (1989); context stories, rules, and instructions were in German.

We generated two random orderings of the 12 problems in each series. Thus, in all, there were four different booklets, with two different series and two different random orderings of problems. For each problem, the order of the four "cards" – that is, the possible responses – was also determined randomly.

Thus each subject received a booklet with instructions and 12 selection tasks. Subjects were instructed to do the problems in order, without going back to a previous problem or changing previous answers. They could take as much time as

Table 1. *Overview*

		Context story/version			
No.	Rule	Series 1	Series 2	Theoretical goal	Empirical test of
1	Cassava	Cheating (SC)	No cheating (no SC)	Replication of Cosmides (1989, Exp. 1)	SC theory against availability theory
2	Duiker	No cheating (no SC)	Cheating (SC)		
3	Overnight	Cheating (SC)	No cheating (SC)	Disentangling "social contract rules" from "cheater-detection algorithm", by means of perspective change	SC theory, PRS theory, availability theory
4	Grover	No cheating (SC)	Cheating (SC)		
5	Winner	No cheating (SC)	Cheating (SC)		
6	Dealer	Cheating (SC)	No cheating (SC)		
7	Day off	Party A[a] (employer)	Party B (employee)	Perspective change in social contract rules with bilateral cheating option	SC theory, PRS theory, availability theory
8	Pension	Party B (employer)	Party A (employee)		
9	Subsidy	Party A (house owner)	Party B (environmental office)		
10	Post office	Original (Party B)	Switched (Party A)	Perspective change in social contract rules with unilateral cheating option	SC theory, PRS theory, availability theory
11	Cholera	Original (Party B)	Switched (Party A)		
12	Drinking	Switched (Party A)	Original (Party B)		

Note: 46 subjects responded to the 12 problems in series 1, and 47 to those in series 2.

[a] The perspectives of party A and party B are defined relative to the cheating options (see Figure 5). From the perspective of party B, *P & not-Q* means that party A cheats.

they wanted. Subjects were tested in small groups and were interrogated after they had completed their tasks to see how they thought they had solved the problems.

Part I: Separating social contract rules from the cheater-detection algorithm

There are two subdivisions in Part I. First, we present our replication of Cosmides' (1989) striking findings concerning unfamiliar social contracts, using students with a different educational background in logic and mathematics. Second, and more important, we experimentally disentangle the theoretical concepts of a "social contract rule" and a "cheater-detection algorithm". The first subdivision provides a test between SC theory and availability theory; the second a test among SC theory, availability theory, and PRS theory.

Unfamiliar social contracts: A replication

According to SC theory, the human mind distinguishes between rules that are social contracts and ones that are not, independent of whether the rules are familiar or not. According to availability theory, the essential distinction is between familiar and unfamiliar rules, independent of whether they are social contracts or not. To test both theories, Cosmides (1989, Exp. 1) used unfamiliar rules, which were surrounded by context stories that cued the rule either as unfamiliar social contracts or as unfamiliar "descriptive" rules (i.e., rules that were not social contracts). Two of these unfamiliar rules were:

(1) If a man eats cassava root, then he must have a tattoo on his face.
(2) If you eat duiker meat, then you have found an ostrich eggshell.

We used the same four context stories as Cosmides (1989, problems 1, 3, 4, and 6, pp. 263–268). As mentioned above, in the context story that cued the cassava rule into a social contract, it was stated that cassava root is an aphrodisiac. The elders have made rule (1) a law in order to ration the rare cassava root for the married. The subjects were cued into the perspective of a person who enforces the law, that is, who checks whether any of four Kaluame men were violating the law (see Figure 1). Since the rule was identified as a social contract (SC), and the subjects were cued into the perspective of a party who can be cheated by the other party, we refer to this version as the *cheating (SC)* version.

The other context story did not cue the rule into a social contract (what Cosmides referred to as a "descriptive" rule). Here, subjects were cued into the perspective of an anthropologist who has been told that rule (1) holds among the Kaluame and who wants to find out whether it actually does. A rationale was suggested for the rule *that completely avoided any reference to a social contract* (i.e., men with tattoos live in a different place from men without tattoos; cassava root grows only where the men with tattoos live). Thus the unfamiliar rule was not identified as a social contract. We refer to this as the *no-cheating (no SC)* version. The subjects again received the same four cards that represented information about four men and were asked to turn over the card(s) that could show if any of these men violated the rule.

For the duiker rule – rule (2) – Cosmides constructed two context stories using the same rationale as above. In the *cheating (SC)* context story, duiker meat was desirable and scarce, and to earn the privilege of eating it a boy must have found an ostrich eggshell, which is a difficult task representing a boy's transition to manhood. This context story cued the rule into a social contract and the subject into an anthropologist who is interested in whether boys ever violate the law, and checks information about four boys. The four cards read: "eats some duiker meat", "has never found an ostrich eggshell", "does not eat any duiker meat", "has found an ostrich eggshell". In the *no-cheating (no SC)* context story, the subject was cued into the role of an anthropologist who wants to find out whether this rule holds. Several explanations were mentioned for the rule (e.g., duikers are small antelopes that feed on ostrich eggs, and they are caught while eating), but *none of them suggested that the rule could be a social contract.*

According to SC theory, if a rule is perceived as a social contract, then a cheater-detection algorithm is activated that searches for information that could detect cheaters. In rules (1) and (2) these are the *P* card and the *not-Q* card. Thus, if SC theory is correct, there will be a high proportion of *P & not-Q* responses in the social contract versions of the rules, but a low percentage in the descriptive version of the same rules. If availability theory, on the other hand, is correct, we expect no difference between the two versions and a low proportion of *P & not-Q* answers in each, since the rules are unfamiliar. For instance, it is difficult to see how the association "eats cassava root" and "has no tattoo on his face", that is, *P & not-Q*, could be available from long-term memory. Neither theory predicts *not-P & Q* responses – a rare response in selection tasks, which will, however, be of crucial importance in Part II.

Out of 93 subjects, 46 answered the cassava problem in its *cheating (SC)* version, and 47 in its *no-cheating (no SC)* version. The latter group answered the duiker problem in the *cheating (SC)* version, and the former group in its *no-cheating (no SC)* version (see Table 1). Predictions and results are shown in Table 2. Averaged across both rules, there were 94% *P & not-Q* answers in the *cheating (SC)* versions, compared with 44% in the *no-cheating (no SC)* versions. Cosmides (1989, Exp. 1) reported 75% and 21%, respectively. *Not-P & Q* responses were very rare.

We have two results. First, we could replicate the effect of a social contract over a non-social contract version – that is, the difference between the two versions – as precisely as one would wish: a 50 percentage point difference in our

Table 2. *Reasoning about unfamiliar rules* (n = *93): Replication of Cosmides (1989, Exp. 1)*

	Predictions		Results (in %)			Cosmides Average
	Social Contract	Availability	Cassava	Duiker	Average	
P & *not-Q* responses						
Cheating (SC)	High	Low	96	91	94	75
No cheating (no SC)	Low	Low	36	52	44	21
not-P & *Q* responses						
Cheating (SC)	Very low	Very low	0	0	0	0
No cheating (no SC)	Very low	Very low	6	0	3	0

Note: Numbers are percentages of subjects choosing *P* & *not-Q* or *not-P* & *Q* responses. SC stands for "social contract". Predictions are from Cosmides (1989). All percentages are rounded.

study as compared with 54 points in Cosmides'. Second, in *both* versions our Konstanz students gave more *P & not-Q* responses (by 20 percentage points) than the Harvard undergraduates participating in Cosmides' study. In fact, 96% *P & not-Q* answers (44 out of 46) in the cassava rule seems to be one of the largest "content effects" ever reported in a selection task. Our Konstanz students seem to be more inclined to reason by a mental logic that follows propositional logic than Cosmides' Harvard students. We can only speculate about the reasons; the rigorous training in mathematics (including proof techniques such as *reductio ad absurdum* that operate by deducing contradictions) in the German gymnasium (pre-university education) is a plausible explanation.

These results give strong support to SC theory in comparison to availability theory. Similarly, they equally strongly support PRS theory in comparison to availability theory, since the social contract versions of both rules are permission rules ("If the action is to be taken, then the precondition must be satisfied"). But our replication also suggests that the degree to which some kind of acquired mental logic supplements "Darwinian algorithms" or "permission schemas" may vary substantially between human subgroups and educational systems.

What could proponents of availability explanations conjecture? It seems to us that they would have to concur that SC theory and PRS theory give the better predictions for unfamiliar rules. But availability theorists would still have a rejoinder: these results do not imply that either of these two theories does better than availability for *familiar* rules. For instance, availability might be the (only) decisive cognitive process in familiar rules, whereas entirely unfamiliar rules activate other processes – such as a cheater-detection algorithm or a permission schema. The hypothesis of two separate processes for familiar and unfamiliar rules parallels Griggs and Cox's (1983) hypothesis that memory-cueing (availability) is activated by familiar rules, whereas unfamiliar rules activate "short-circuiting processes" such as "matching bias" and "verification bias".

This conjecture is not without foundation. Cosmides (1989) did not compare social contract and "descriptive" versions of *familiar rules* (but see Cosmides, 1985, p. 244–251). Cheng and Holyoak did attempt such a comparison (1985, Exp. 1), but one that has been criticized for conflating availability with permission schemas. Cheng and Holyoak added context information to a rule (in order to cue a permission schema) and compared it with the same rule without the context added. An availability theorist, Cosmides (1989, p. 204) argued, could claim that adding a context to a rule can cue many additional memories, and increase the probability that a falsifying response (*P & not-Q*) would have become available from long-term memory. Therefore Cheng and Holyoak's finding that the proportion of *P & not-Q* responses increased when the context was added supports, on Cosmides' argument, both PRS theory *and* availability theory. The issue is far from having an easy answer, if only because the concept of availability and the formulations of the memory-cueing view are notoriously unclear. Nevertheless, the argument still holds that proponents of SC theory and PRS theory have not yet shown that availability theory fails to predict performance in familiar social contract rules.

We addressed this argument in the eight problems tested in the following section, which allow for a test of SC theory against availability that uses *familiar* rules, but avoids the potential confounding from adding extra context information. Unlike Cosmides (1989), we compared familiar rules in both context story versions. Like Cosmides, but unlike Cheng and Holyoak (1985), we constructed all context stories to be about the same length (the same as the "cassava" and "duiker" context stories). And, unlike Cosmides (1989), we gave each pair of context stories similar content. Since testing this defence of availability theory was not the main purpose of the following section, we want to mention the result without further ado: with familiar rules, SC theory gives better predictions than availability theory, too (see Figures 2 and 3). Thus availability theory cannot be saved by making the dual processes distinction.

Are social contract rules sufficient for P & not-Q *responses?*

We now want to disentangle two theoretical concepts in SC theory. The first is the notion that a rule is a social contract; the second is that of a cheater-detection algorithm.

What is the relationship between these two central concepts in SC theory? Is the fact that a rule (such as the cassava rule) is perceived as a social contract a sufficient condition for the *P & not-Q* responses observed in the preceding section? Or is it necessary that in addition a cheater-detection algorithm be activated? Can we separate the two conceptions at all? Cosmides seems not to have attempted a sharp distinction between these two concepts. For instance, we understand her as saying that for the social contract algorithms to operate there must be both a rule having the characteristic structure of a social contract and a cheater-detection algorithm (1989, pp. 223–230). But elsewhere, one can understand her as saying that only the former is sufficient: "Thus, for social contract theory, the major determinant of responses is whether a rule is a social contract (SC) or descriptive" (p. 207).

In all her tests, social contract rules and cheater-detection algorithms went together, so there was no pressure to distinguish conceptually between them. In her tests of SC theory against availability she cued a rule as either a social contract or a "descriptive" rule, and performance in all social contract rules was explained by a cheater-detection algorithm. Similarly, in all tests of SC theory against PRS theory, she cued a rule as either a social contract or a conditional permission that was not a social contract, and performance in all social contracts was attributed to the cheater-detection algorithm. Thus one could easily read the experiments as saying: if a rule is cued as a social contract, then a cheater-detection algorithm is activated in the subject's mind.

But what is the crucial cognitive process? That a rule is understood as a social contract? Or that the cheater-detection algorithm is activated? We have designed eight problems (four rules, two context stories each) that can disentangle the two theoretical notions. Thus the first purpose of the following test is to gain better insight into the nature of the cognitive processes postulated in social contract theory – by disentangling "cheater-detection" from "perceiving a rule as a social contract". The second purpose is to provide a sharp test of the concept of a cheater-detection algorithm, which is at the very heart of SC theory – and thereby to provide a test between SC theory and PRS theory.

Recall that in her descriptive context stories Cosmides avoided all cues that might indicate even the possibility that the rule was a social contract. Rather, she was at pains to construct a story with a non-social contract rule (e.g., that tattooed men happened to live where cassava roots grow). (Similarly, Cheng & Holyoak, 1985, Exp. 1, avoided identifying the rule in their "no-rationale" version as a permission rule.) Thus cheater-detection algorithms and social contracts always went together. We will now test SC theory in problems where *in both versions the rule was identified as (the same) social contract*, but being cheated was possible in only one version.

We used two social contract rules that were familiar permission rules (in Cheng & Holyoak's terminology), and two others that were familiar obligation rules.

Permission rules

The two permission rules were:

(3) If someone stays overnight in the cabin, then that person must bring along a bundle of wood from the valley.

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

The overnight rule – rule (3) – is familiar to generations of mountain hikers in the Alps. Two context stories were used. The first explained that there is a cabin at high altitude in the Swiss Alps, which serves hikers as an overnight shelter. Since it is cold and firewood is not otherwise available at that altitude, the rule is that each hiker who stays overnight has to carry along his/her share of wood. There are rumours that the rule is not always followed. The subjects were cued into the perspective of a guard who checks whether any one of four hikers has violated the rule. The four hikers were represented by four cards that read "stays overnight in the cabin", "carried no wood", "carried wood", and "does not stay overnight in the cabin". We refer to this version of the overnight problem as the *cheating version*.

In the *no-cheating version* the subjects were cued into the perspective of a member of the German Alpine Association who visits the Swiss cabin and tries to find out how the local Swiss Alpine Club runs this cabin. He observes people bringing wood to the cabin, and a friend suggests the familiar overnight rule as an explanation. The context story also mentions an alternative explanation: rather than the hikers, the members of the Swiss Alpine Club, who do not stay overnight, might carry the wood. The task of the subject was to check four persons (the same four cards) in order to find out whether anyone had violated the overnight rule suggested by the friend.

As mentioned above, in both versions of the overnight rule the four cards and the instructions were the same. This holds for all 12 rules. The instruction always followed the violation format used by Cosmides: "Indicate only the card(s) you definitely need to turn over to see if any of these [hikers, Kaluame men, etc.] violate the rule."

Note that in both context stories the rule was explicitly identified as the same social contract. The difference was that in the cheating version the subject was cued into the perspective of one party in a social contract, and the other party (the hikers) had a cheating option. In the no-cheating version, the subject was cued into the perspective of a third party who is not engaged in the social contract, but who attempts to determine whether the social contract rule exists. From the perspective of the local guard, a violation of the rule indicates *cheating*, whereas from the perspective of the member of the German Alpine Association, a violation indicates that the proposed rule is *incorrect*. The social contract rule is seen either as deliberately or negligently violated by a human, or as descriptively wrong.

The cheater-detection algorithm in SC theory predicts a high proportion of *P & not-Q* responses when cheating is possible, and a low percentage otherwise. Availability theory, in contrast, predicts no difference, because the rules are identical and the context stories are similar in content and length. PRS theory also predicts no difference between the two versions, because the rule is in both cases a permission rule and cheating options and perspectives do not enter into PRS theory. PRS theory should predict for both versions a high percentage of *P & not-Q* responses. Predictions are shown in Figure 2.

We used the same rationale for the "Grover High School" rule. This rule has been studied before, with mixed results. For instance, van Duyne (1974) found a substantial "content effect" (*P & not-Q* responses), whereas Yachanin and Tweney (1982) did not. For the cheating version, we used Cosmides' (1989, problem 9, pp. 269–270) social contract context story, which states that parents in Grover City spend a lot of money on Grover High School, whereas parents in Hanover, although equally prosperous, never wanted to spend much money on Hanover High School. Consequently, children at Grover High School get a better education. This was the reason why the Board of Education set up rule (4). The subjects were cued into the perspective of a supervisor at the local Board of

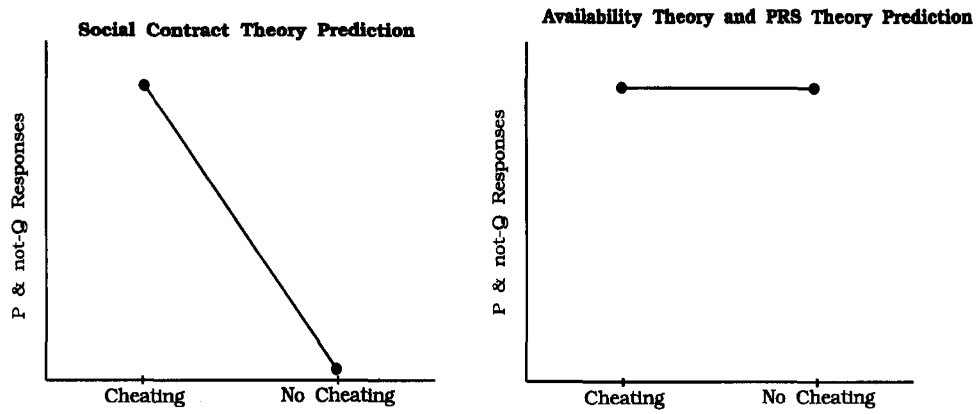


Figure 2. *Predictions for rules (3) to (6). All rules are social contracts. SC theory predicts a decrease of* P & not-Q *responses when subjects are cued into a perspective in which they cannot be cheated. Availability theory and PRS theory predict no difference.*

Education, who checks whether any one of four volunteers to the Board of Education – who assigned their own teenagers to the two high schools – has violated the rule. Each card represented the child of one volunteer. One side told what school the volunteer assigned her child to, and the other side told what town the student lived in.

In the no-cheating version the subject was cued into the perspective of a representative of the German government who visits the USA. The representative is informed that in the USA the quality of a school depends heavily on the willingness of a community to spend money on schools, and the case of Grover High School and Hanover High School is presented as an example. As in the cheating version, the context story specifies that the parents in Grover City have always cared about the quality of their schools, including Grover High, and have been willing to pay for it. In contrast, the parents in the neighbouring town of Hanover have never wanted to spend the money. Therefore the quality of Grover High is much better than Hanover High. The representative wonders how students are assigned to the two schools, given the different willingness to pay for education. A colleague suggests rule (4) as a plausible assignment rule created by the local Board of Education, given that situation. The context story also mentions two other possible assignment rules: assignment by the student's ability and achievement only; and assignment by the spatial distance between a student's home and a school. The context story said that the representative is interested in checking the validity of the rule suggested. Again, the four cards represented information about four students. The subjects' instructions were the same as in the cheating version: to indicate only those card(s) you definitely need to turn over to see if any of these students violate the rule.

As before, in both versions the Grover rule was explicitly identified as the same social contract, with the same cost and benefit structure. Predictions are the same as for the overnight rule and shown in Figure 2.

Figure 3 shows that the results are similar for both social contract (permission) rules. The percentages of *P & not-Q* responses were 89% ("cheating" version) and 53% ("no cheating") for the overnight rule, and 77% and 46% for the Grover rule. On the average, the proportion of *P & not-Q* responses was 33 percentage points higher in the cheating version than in the no-cheating version. These results support the concept of a cheater-detection algorithm, as postulated by SC theory. The key concepts of a social contract rule and a cheater-detection algorithm can be experimentally separated. The decisive concept is the cheater-detection algorithm. The fact that a rule is perceived as a social contract is not sufficient by itself for the striking results reported by Cosmides (1989, Exp. 1) and replicated above. Availability theory cannot account for these results. Nor can PRS theory: in both versions the rule is a permission rule and "cheating" per se plays no role in PRS theory. PRS theory predicts subjects will be good at detecting violations of permission rules, but is mute on whether the violations represent cheating or not.

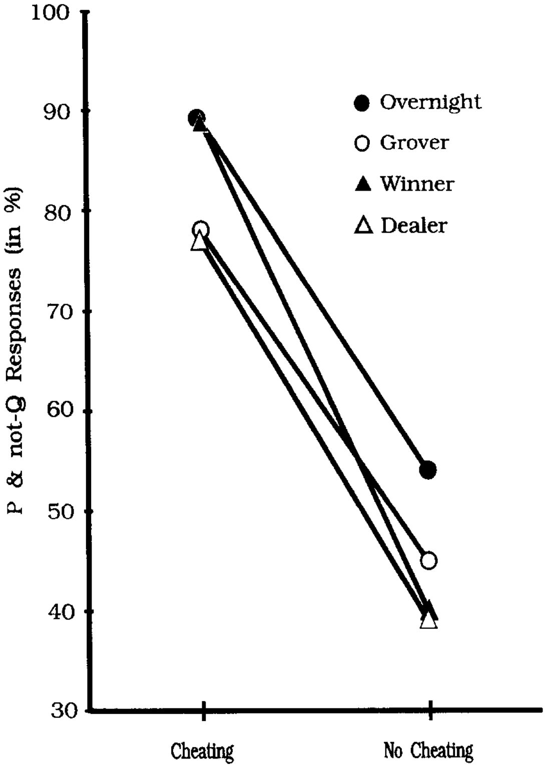


Figure 3. *Results for the Overnight and Grover rules (permission rules) and the Winner and Dealer rules (obligation rules). All rules are social contracts.*

Again, the overall percentage of *P & not-Q* responses was very high, particularly in the overnight rule.

Obligation rules

We will now submit the cheater-detection algorithm in SC theory to the same strong test using two familiar rules that are obligations rather than permissions (in Cheng & Holyoak's terminology):

(5) If a player wins a game, then he will have to treat the others to a round of drinks at the club's restaurant.
(6) If a small-time drug dealer confesses, then he will have to be released.

Both rules have the structure of obligations: if a precondition is satisfied, an action must be taken. The winner rule – rule (5) – is a rule used by Cheng, Holyoak, Nisbett, and Oliver (1986), and was slightly modified to sound familiar to our German subjects. We added a cheating context story in which the subject was cued into the perspective of a member of a tennis club, whose members play tournaments and must follow the winner rule if they win. There are rumours that the rule has been violated before. The subject had to check whether any of four players violated the rule. The four cards read: "won the game", "did not pay for a round of drinks", "did not win the game", "paid for a round of drinks". In the no-cheating context story, the subject was cued into the role of a visitor to the tennis club, who observes that several people treat the others to a round of drinks. It is explicitly stated that they may not be doing this voluntarily, but rather following the familiar obligation rule (5). The context story also mentions a possible alternative explanation: that the players in the club are simply generous and friendly. The visitor's task was to check information on four players to find out whether any information violated the winner rule.

In the cheating version of the dealer rule, subjects were cued into the role of a small-time drug dealer arrested by the police and promised that he would be released if he confessed and provided information about the "big sharks". Confession and information is a benefit for the police, but a cost for the dealer, who must anticipate revenge from the big sharks. Being released, on the other hand, is a benefit for the dealer. The subject's task was to check information about four similar cases that could reveal that the police had violated the rule before. The four cards read: "confessed", "was not released", "did not confess", "was released". In the no-cheating version, the subject was cued into the role of a journalist who is investigating police methods in dealing with drug crimes. The rule was again identified as a social contract that the police might possibly use. The journalist was interested in finding out whether the police indeed used this unconventional social contract.

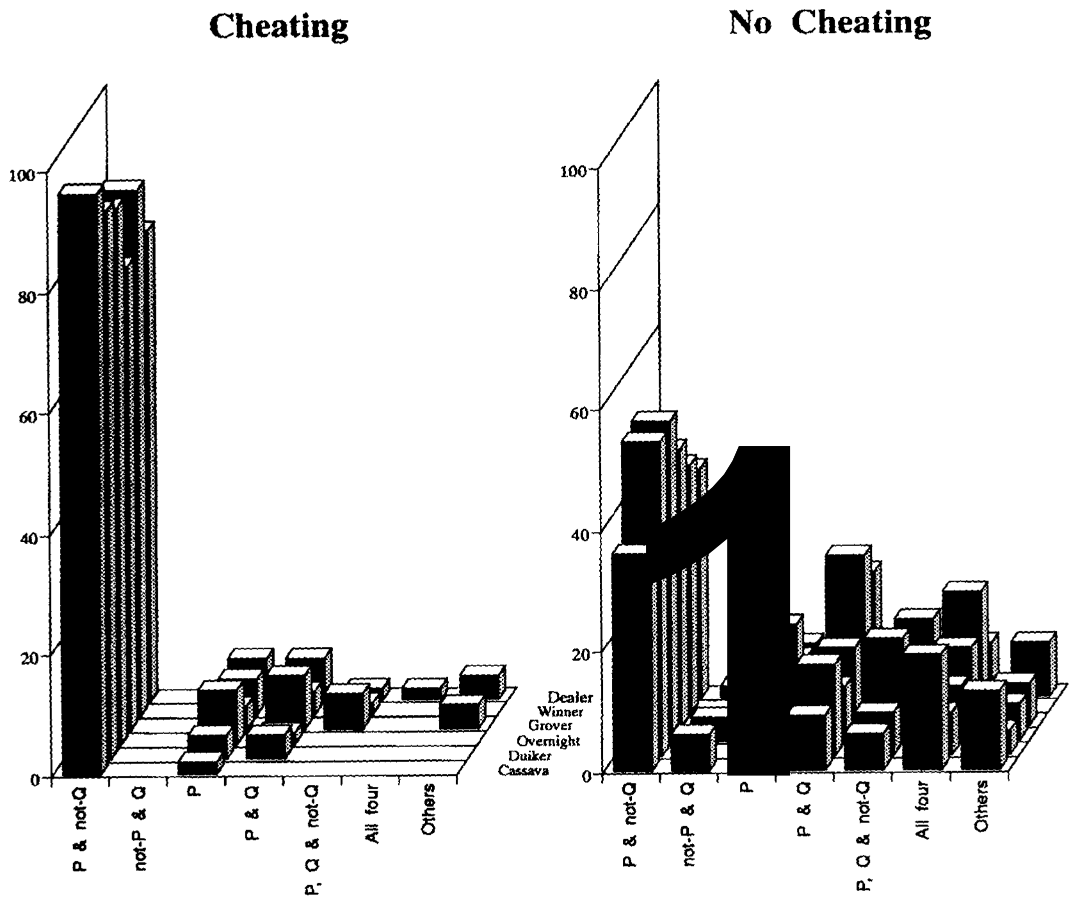
Figure 3 shows the results for both obligation rules. The percentages of *P & not-Q* responses were 89% and 41% for the winner rule, and 77% and 38% for the dealer rule. The average difference in *P & not-Q* responses was 44 percentage points, even larger than for the permission rules. These results support the concept of a cheater-detection algorithm in SC theory, and are inconsistent with availability theory. PRS theory cannot account for the differences found: the rule in both versions is an obligation rule; thus there should be no difference.

The fact that a rule is perceived as a social contract, a conditional permission or obligation does not by itself explain Cosmides' (1989) results and the responses of our subjects. *In fact, rules which were not social contracts (no SC) elicited virtually the same average percentage of P & not-Q responses (44%, see Table 2) as social contract rules (45%, on the average, see Figure 3) when subjects were cued into a perspective that does not activate a cheater-detection algorithm.*

Response distribution

Figure 4 shows the distribution of the responses for all six rules (12 problems). There are 16 possible selections from four cards. In social contracts with cheating options by the other party, residual responses were few and predominantly *P* or *P & Q*. The distribution in the no-cheating versions, in contrast, was comparatively flat. It included substantial proportions of *P*; *P & Q*; *P, Q & not-Q*; and "all four" responses. Note that all frequent responses included the *P* card. The *not-P & Q* response was extremely rare. It never occurred in the cheating versions of rules (1) to (6). In the no-cheating versions, the average percentage of *not-P & Q* responses was 3% in Table 2, 2% for the permission rules (3) and (4), and 1% for the obligation rules (5) and (6). In Part II we explain how we cued subjects into perspectives that predict this rare response.

What conclusions about the performance of individual subjects can we draw from these two response distributions? Do the distributions reflect systematic individual differences between persons (in particular, in the no-cheating problems), or does the variation in the aggregate mirror intra-individual response variation? Because we had every person answer three cheating and three no-cheating problems, there are data to answer this question. We counted all cases where a subject gave the same response three times in the three problems with cheating options, and we did the same for problems without cheating options. We refer to three identical responses as a "consistent pattern". If the response distributions were due only to systematic individual differences, then we should expect no difference in the percentages of consistent patterns between the cheating and no-cheating versions, and they should be close to 100%. The actual percentages of consistent patterns were, however, 67% and 28%, for the cheating and no-cheating versions, respectively. All consistent patterns in the cheating versions and 18 out of the 26 consistent patterns in the no-cheating versions were *P & not-Q* responses. These findings indicate that neither the few residual answers in the cheating versions nor the "flat" distribution in the no-cheating versions can be accounted for by systematic individual differences in responses.



Cheating

No Cheating

Figure 4. *Response distribution for rules (1) to (6). Responses are shown in percentages. Left: Responses on problem versions with cheating options. Right: Responses on versions without cheating options.*

The transition from small to large variability in Figure 4 seems to reflect largely a transition within the response pattern of individual subjects.

Summary and discussion: Part I

So far, we have shown two things. First, we replicated Cosmides' (1989, Exp. 1) tests of SC theory against availability theory, using unfamiliar rules that were either perceived as social contracts or not. The difference between the proportion of subjects selecting *P & not-Q* in the two versions was 50%, almost identical with what Cosmides reported. This strong effect contradicts availability theory, but gives support to SC theory. We also showed that our Konstanz students' judgments more often followed propositional logic than those of Cosmides' Harvard students, independent of whether the rule was perceived as a social contract or not.

Second, we experimentally separated the concept of a social contract rule from the concept of a cheater-detection algorithm. The main theoretical result from this is: that a rule is perceived as a social contract rule is *not* a sufficient condition for obtaining Cosmides' results – contrary to what she seems to claim in her 1989 article, but consistent with the conceptual framework of SC theory. In fact, the average proportion of *P & not-Q* responses was *almost identical* for rules not perceived as social contracts (cassava, duiker) and rules perceived as social contracts (all others), unless the subject was cued into the perspective of a party that could be cheated. Thus, whether a rule was perceived as a social contract or not made *no* difference by itself. The decisive theoretical concept seems to be the cheater-detection algorithm.

The challenge our results pose for PRS theory is to provide an independent motivation for accounting for the cheating option, because, in contrast to SC theory, permission and obligation schemas do not invoke cheater-detection algorithms as a concept.

What is at issue is not simply the "semantic" question whether a rule is a social exchange, pseudo-exchange, or a permission or obligation rule, as it seemed from the previous debate (Cheng & Holyoak, 1989; Cosmides, 1989; Pollard, 1990). Reasoning about the six conditionals above is more pragmatic than previously supposed: a definition of the domain "social contracts" or "social regulations" (permissions, obligations) must include the perspective of the subject – as a participating party who can be cheated, or as a disinterested third party who determines whether or not a proposed social contract rule actually holds.

"Instructional effects"

The present results can help one to understand a body of apparently inconsistent results on so-called instructional effects. In one group of studies, exemplified by Wason's (1966) original study, the instruction read "determine if the rule is true or false"; in another group, subjects were asked to "determine if the rule is being violated". Yachanin and Tweney (1982) argued that the "facilitation" (large proportion of *P & not-Q* responses) reported in earlier experiments was mainly a consequence of using violation instructions. The explanation proposed was this: in the true–false version, where one reasons *about* a rule, the truth of the rule has to be assessed and the subject has to consider *two* hypotheses: that the rule is true and that the rule is false. In the violation version, where one reasons *from* a rule, potential violaters have to be identified. In the case of two hypotheses, the "cognitive load" would be greater and might lead more frequently to "cognitive short-circuiting strategies" such as confirmation and matching bias. Griggs (1984) used one familiar rule (drinking rule, see below) and one unfamiliar rule (an abstract rule), and true–false and violation instructions of each. He concluded, however, that violation instructions were neither necessary nor sufficient for facilitation. Yachanin (1986) used the same materials and obtained the same results. Valentine (1985) could also not support Yachanin and Tweney's conclusion; she found no difference between instructions. These authors all discussed their respective results using concepts such as "familiarity" (availability), "memory cueing", "cognitive load" and "short-circuiting strategies".

There is another way to understand these "instructional effects" that seem to appear and disappear. We can connect the two kinds of instructions – and the associated purposes of assessing the truth of a rule versus detecting violations – with the concepts of perspective and cheating detection.

According to the present, revised prediction of SC theory (where social contracts and the cheater-detection algorithm are disentangled through the introduction of perspectives), no general effect of the true–false versus violation instructions is expected – in contrast to Yachanin and Tweney's (1982) proposal. Rather, a large proportion of *P & not-Q* responses is predicted if (1) the rule is perceived as a social contract and (2) the subject is cued into the perspective of a party that can be cheated. This is exactly the violation instruction where Griggs (1984) finds 100% ($n = 25$) *P & not-Q* responses in the drinking rule. If only (1) holds, he reports only 64%. Yachanin (1986) replicated this difference: 85% and 50% ($n = 20$), respectively, and so did we (see Table 3, below). If the rule is, however, an abstract rule (no social contract, permission or obligation) such as "if there is an 'A' on one side, then it must have a '3' on the other side", then the two instructions do *not* cue different perspectives as they do with social contracts, nor is a cheater-detection algorithm involved. Therefore, in abstract rules the two instructions should make little difference. This is what Griggs (1984), Valentine (1985) and Yachanin (1986) found, but they seemed to have no explanation except that "some familiarity with the rule is necessary if violation instructions are to have any facilitating influence" (Yachanin, 1986, p. 26). We argue that the theoretical concept needed to understand this apparent inconsistency is not familiarity (availability), but the concepts of perspective and cheating option. Moreover, Table 2 shows that familiarity is *not* a necessary condition for violation instructions to produce large proportions of *P & not-Q* responses.

Availability

Can we find an effect of availability at all in Table 2 and Figure 3? Note that even if availability cannot account for the large effect due to the cheating option, it should predict that *P & not-Q* responses are, overall, more frequent in the familiar rules. But this is not the case: their overall percentage (cheating and no-cheating versions) is in fact larger in unfamiliar rules (Table 2) than in familiar rules (Figure 3). Thus for rules (1) to (6) *we cannot even find an effect of availability that adds to the large effect predicted by SC theory.*

Throughout Part I, a change in perspective was coupled with a change in purpose of reasoning. By cueing a subject into the perspective of a party that can be cheated, reasoning about "violations" was tuned towards the purpose of detecting cheating. By cueing the subject into the perspective of a third party not engaged in the social contract, reasoning was directed towards checking whether a social contract actually exists. In Part II we test situations where a change in perspective is *not* coupled with a change in purpose.

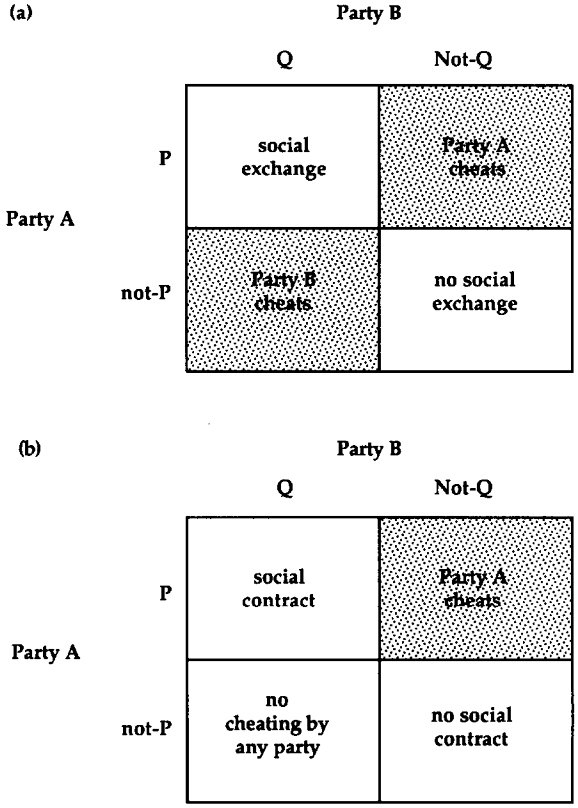
Part II: Cheating and perspective change

A social contract engages two parties, be they individuals, social groups, or institutions. It is not always the case that each party in a contract can cheat the other. Consider, for instance, the following postal rule currently valid in Germany: "If an envelope is sealed, then it must have a 1-mark stamp." The parties are the sender and the post office. The sender can try to cheat by putting a 60-pfennig stamp (the postage for an unsealed envelope) on a sealed envelope (*P & not-Q*); the post office, however, cannot cheat on this rule. *Not-P & Q* does not count as cheating. If the sender chooses *not-P & Q*, that is, puts a 1-mark stamp on an unsealed envelope, she causes unnecessary costs for herself – a kind of self-cheating, at best. Overpayment may indicate carelessness or that she wants to subsidize the post office. But it does not mean that the post office has cheated. (Cheating by the post office would be deliberately failing to deliver letters with proper postage, which is not an option in the context of the rule.)

We refer to social contracts where only one party can actively cheat as social contracts with a *unilateral cheating option*. One reason for unilateral cheating in the post office rule is that *P* and *Q* are both actions performed by the same party; but there are other reasons for unilateral cheating. We will return to these shortly.

We can define social contracts with a *bilateral cheating option* as follows: from the perspective of party A, *not-P & Q* means that party B cheats party A; from the perspective of party B, *P & not-Q* means that party A cheats party B (see Figure 5a). This definition presupposes (1) that *P* and *not-P* are actions of party B, and *Q* and *not-Q* are actions of party A, and (2) that *P* is a benefit for party A and a cost for party B, and *Q* is a cost for party A and benefit for party B. We will also refer to a social contract with such bilateral cheating options as "social exchange", to distinguish it from other social contracts where this symmetry does not hold.

Most of the previous research on the selection task has investigated social contracts with unilateral cheating options (Figure 5b), and the subjects were cued into the perspective of party B. The cheating versions of the six rules we used in Part I had, with one exception, only unilateral cheating options.[5] Neither the rules nor their contexts invited an interpretation of the social contract as involving



(a)

		Party B	
		Q	Not-Q
Party A	P	social exchange	Party A cheats
	not-P	Party B cheats	no social exchange

(b)

		Party B	
		Q	Not-Q
Party A	P	social contract	Party A cheats
	not-P	no cheating by any party	no social contract

Figure 5. *Social contracts with (a) bilateral and (b) unilateral cheating options.*

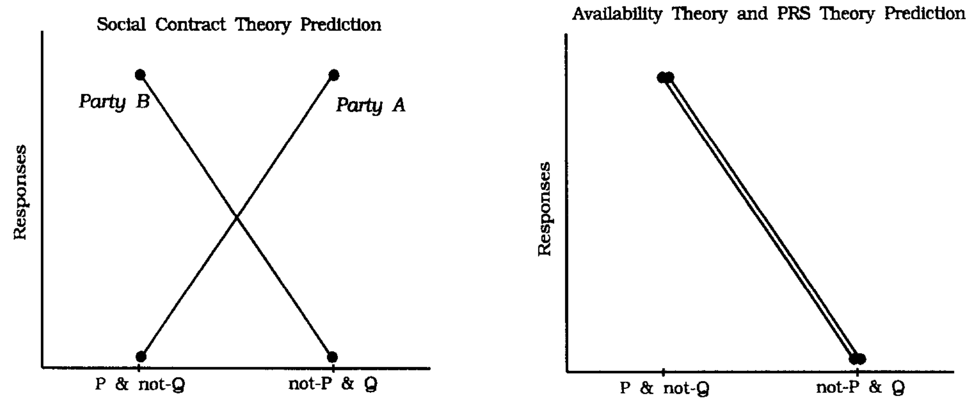
[5] The exception is the dealer rule, where a faked confession by a drug dealer and the dealer's subsequent release (*not-P & Q*) can be seen as an instance of cheating by the drug dealer (party A).

bilateral cheating options. In the overnight problem, for instance, *not-P & Q* means that a hiker brought a bundle of wood, but did not stay overnight. This outcome can mean that unexpected events interfered that kept the hiker from staying overnight. As in the case of the post office rule, it is not, however, an instance of cheating by the other party.

Perspective change in social contracts with bilateral cheating options

Symmetrical cheating options allow for a novel test of SC theory. We switched the subjects' perspective (from party B to party A), but held the rule constant – both in its actual wording and in its interpretation as a particular social contract in the context story. SC theory postulates that a cheater-detection algorithm is activated if a subject represents one party in a social contract and the other party has a cheating option. If this is correct, the following responses must be observed. If subjects are cued into the perspective of party B, a cheater-detection algorithm predicts a high percentage of *P & not-Q* responses – as in rules (1) to (6). If subjects are cued into the perspective of party A, the same algorithm predicts a similarly high perçentage of *not-P & Q* responses (see Figure 6).

What do competing theories predict for a perspective change? No availability theory seems to have ever predicted *not-P & Q* responses. Since all the following rules are familiar to our subjects, availability theory should predict a mid-to-high percentage of *P & not-Q* responses, and, most important, no difference between perspectives. What does PRS theory predict? The concept of a cheater-detection algorithm is not part of the production rules that define a permission/obligation schema. Nor are the concepts of two parties and two perspectives involved, which are intimately tied to the cheater-detection algorithm. PRS theory has no



Social Contract Theory Prediction

Availability Theory and PRS Theory Prediction

Figure 6. *Perspective change. Predictions for rules (7) to (9). All rules are social contracts. Party B can be cheated by* P & not-Q; *party A can be cheated by* not-P & Q *(see Figure 5a).*

theoretical vocabulary that applies to perspective change. Since all rules used are permission/obligation rules, and do not change their status when the perspective is changed, PRS theory should also predict a high proportion of *P & not-Q* responses and low proportions of *not-P & Q* responses under both perspectives (see Figure 6).

We used three social exchange rules, that is, social contracts with bilateral cheating options. The first was the "day off" rule:

(7) If an employee works on the weekend, then that person gets a day off during the week.

Two context stories were used. One of them cued the subject into the employee's perspective, the other into the employer's. The employee version stated that working on the weekend is a benefit for the employer, because the firm can make use of its machines and be more flexible. Working on the weekend, on the other hand, is a cost for the employee. The context story was about an employee who had never worked on the weekend before, but who is considering working on Saturdays from time to time, since having a day off during the week is a benefit that outweighs the costs of working on Saturday. There are rumours that the rule has been violated before. The subjects' task was to check information about four colleagues to see whether the rule has been violated before. The four cards read: "worked on the weekend", "did not get a day off", "did not work on the weekend", "did get a day off".

In the employer version, the same rationale was given. The subject was cued into the perspective of the employer, who suspects that the rule has been violated before. The subjects' task was the same as in the other perspective.

Thus from the perspective of the employee, the employer would have cheated when an employee worked on the weekend but did not get a weekday off (*P & not-Q*). From the perspective of the employer, the employee would have cheated when he did not work on the weekend but took a weekday off anyway (*not-P & Q*).

How were judgments affected by perspective change? The dominant *P & not-Q* response (75%) change into a *not-P & Q* response (61%), as predicted by SC theory. The infrequent *not-P & Q* response (2%) changed into an infrequent *P & not-Q* response (15%). Note that the differences between the two perspectives *were equally large* (60 and 59 percentage points) for both *P & not-Q* and *not-P & Q* responses. These figures reveal a tendency in both perspectives to give more *P & not-Q* than *not-P & Q* responses, which is in part an artifact due to a small group of six "mental logic" subjects who throughout all selection tasks looked for the *P & not-Q* information.[6] Figure 7 illustrates the results excluding

[6] We designate this the "mental logic" group because all six subjects, when interviewed afterwards about the procedures they had used to answer the questions, reported that they had deliberately applied the rules of the standard truth table.

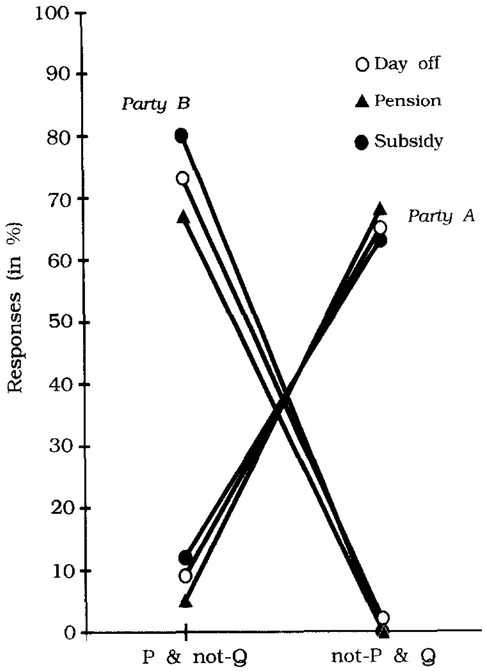


Figure 7. *Perspective change. Results for the Day off, Pension, and Subsidy rules. Note:* n = *87 (the six "mental logic" subjects are not included; see text).*

subjects (percentages differ only slightly). Results of the "day off" rule support the notion of a cheater-detection algorithm, an algorithm controlled by the perspective a person is cued in.

The second social exchange rule we used was the "pension" rule:

(8) If a previous employee gets a pension from a firm, then that person must have worked for the firm for at least ten years.

Context stories were constructed in the same way as for the "day off" rule. Working for at least ten years was described as a benefit for the firm, but as a cost for the employee, who was said to be not entirely happy with the firm. Getting a pension is a benefit for the employee, but a cost for the firm. The version that cued the subject into the perspective of the employee stated that the employee had heard rumours that the rule had been violated previously. So did the version in which the subject was cued into the perspective of the employer (firm). Again, for both parties the task was to check four employees to find out whether the rule has been violated. The four cards read: "got a pension", "worked ten years for the firm", "did not get a pension", "worked eight years for the firm". Figure 7 shows that predictions of SC theory were again confirmed. The percentages of subjects choosing *P & not-Q* were 70% and 11% for the two perspectives, respectively; percentages of *not-P & Q* were 0% and 64% ($n = 93$). The difference between perspectives is again about the same for both *P & not-Q* and *not-P & Q* responses, and of similar magnitude as in the "day off" rule.

The third social exchange rule was the "subsidy" rule:

(9) If a home owner gets a subsidy, then that person must have installed a modern heating system.

In both context stories we explained that the rule was created by an environmental office that is concerned about the pollution caused by out-of-date heating systems in private households. (This is a well-known and serious environmental problem in parts of what was formerly East Germany.) The environmental office offers monetary subsidies to home owners who are willing to install a modern heating system. A subsidy is a benefit for the home owner, but a cost for the environmental office. A modern heating system, in contrast, is a benefit from the perspective of the environmental office, but causes costs (that go beyond the subsidy) for the home owner. The subjects were cued into the perspective of either an environmental officer or a home owner. Both have heard rumours that the rule has been violated before, and check information about four home owners to find out whether the rule has in fact been violated. The four cards read: "got a subsidy", "did not install a modern heating system", "did not get a subsidy", "installed a modern heating system".

Results, shown in Figure 7, again supported SC theory. The percentages of subjects choosing *P & not-Q* were 81% and 17%, and the percentages of *not-P & Q* responses were 59% and 0%, for the two perspectives, respectively ($n = 93$). Note that the size of the differences between perspectives replicated very well over the three rules.

Response distribution

Figure 8 shows the distribution of responses for each perspective ($n = 93$). Seventy-six percent of all responses were *P & not-Q* or *not-P & Q*, for both perspectives. No systematic pattern seems to exist in the residual responses.

Three implications

The prediction and empirical demonstration of *not-P & Q* responses are important for several reasons. First, they rule out one alternative interpretation of the results in Part I as well as of the results reported by Cheng and Holyoak (1985) and Cosmides (1989): namely, that social contracts (permissions and obligations) somehow facilitate logical reasoning. The present result, that perspective change switches a person's response from *P & not-Q* to *not-P & Q*, eliminates this alternative explanation.

Second, and related to this, perspective change illustrates that responses other than *P & not-Q* need not be seen as reasoning errors, but can reflect important

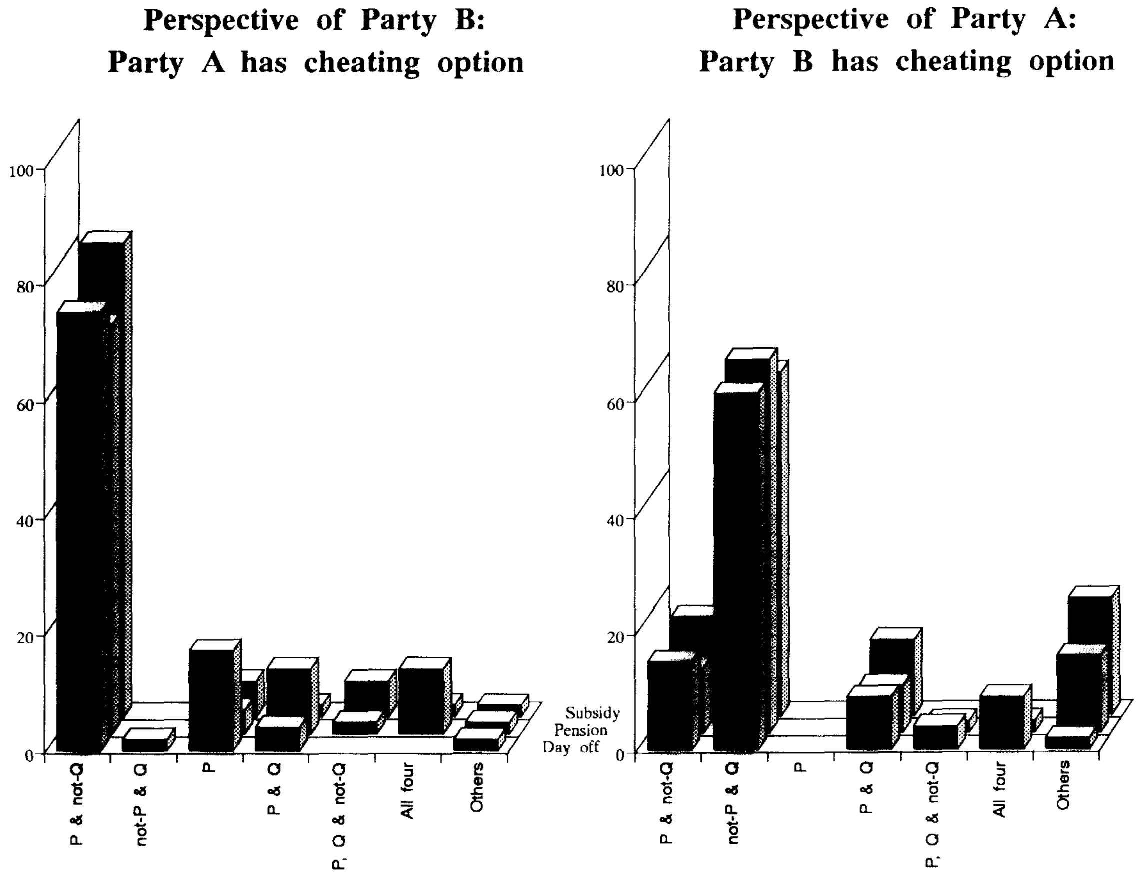


Figure 8. *Response distributions for social contracts with bilateral cheating options.*

adaptive functions. In contrast, earlier theories evaluated such responses merely as deviations from propositional logic, and attributed them to deficiencies in cognitive processes (e.g., verification bias, matching bias) or arbitrary associations (availability).

Third, Cosmides (1989) also predicted and obtained *not-P & Q* responses. She derived these predictions in a different way: by switching the rules from "if *P* then *Q*" to "if *Q* then *P*". Some of her rules, however, such as the cassava rule, could not be switched like abstract rules, and she was forced to change their wording. For instance, the switched version of "If a man eats cassava root, then he must have a tattoo on his face" read "If a man has a tattoo on his face, then he eats cassava root" (p. 217). The deontic "must" in the original rule disappeared. Cosmides stated that this deletion was unavoidable, since a switched rule that keeps the "must" would make little sense ("If a man must have a tattoo on his face, then he eats cassava root"). It seems evident that the switched rule is not quite the converse of the original one; it is less clear what precisely follows from this for evaluating her results (for a critique and discussion see Manktelow & Over, 1987, pp. 207–210; Pollard, 1990). Our predictions of *not-P & Q* responses from perspective change avoided this problem with switching rules. We switched perspectives instead of rules. The results strongly support SC theory.

Perspective change in social contracts with unilateral cheating options

We used three social contract rules with unilateral cheating options: the post office rule, the cholera rule, and the drinking rule. These rules have been tested many times. The post office rule we used was adapted to the current situation in Germany:

(10) If an envelope is sealed, then it must have a 1-mark stamp.

This rule was well known to our subjects. Similar to rules (7) to (9), we used two context stories, each of which cued the perspective of one of the two parties involved in this social contract. One context story cued the subjects into the perspective of a postal official whose task is to check whether any of four letters violates the rule. We refer to this as the "original perspective", because this is the perspective into which subjects have been cued in previous research. The second context story cued the subjects into the perspective of the sender. The sender was an employee of a large firm whose task was to check the outgoing mail. The context story indicated that the other employees of the firm who post letters might sometimes be careless and thereby violate the rule. Again, the wording of the rule, the four cards, and the instruction were kept constant for the two perspectives. The four cards read: "envelope sealed", "60-pfennig stamp", "envelope not sealed", "1-mark stamp".

What is the cost–benefit structure of the post office problem? First, unlike the social exchange rules (7) to (9), both actions (*P* and *Q*) are actions of the same party (the sender or firm). The post office does not perform any of the actions *P* and *Q*. Second, sealing an envelope may be a benefit for the sender, but does not seem to cause any costs for the post office. Here we have an asymmetry in the cost–benefit relation. (Using a 1-mark stamp – as opposed to a 60-pfennig stamp – for an unsealed envelope is a cost for the sender and a benefit for the post office. Therefore with respect to *Q* there is no asymmetry.) Cheating is possible only for the sender, not for the post office.

What does SC theory predict? In the *original* perspective (post office), *P & not-Q* implies that the sender cheats, and therefore SC theory predicts a high percentage of *P & not-Q* responses (and a very low percentage of *not-P & Q* responses). In the *switched* perspective (the sender's), however, the situation is different from social contract rules with a bilateral cheating option: *Not-P & Q* does not mean that the post office is cheating; rather, it may indicate carelessness and unnecessary costs caused by the sender. In the switched perspective, the post office cannot cheat the sender. SC theory should now predict a low percentage of *P & not-Q* responses, because a cheater-detection algorithm is not activated. For the same reason, SC theory should also predict a very low percentage of *not-P & Q* responses (*not-P & Q* indicates no cheating by any party – just carelessness or foolishness on the part of the sender).

What does availability theory predict? Manktelow and Evans (1979) argued that the high percentage of *P & not-Q* responses in the post office rule reported by Johnson-Laird, Legrenzi, and Sonino Legrenzi (1972) for British subjects may have been due to long-term memory cues. A post office rule existed in Britain, and violations (a sealed but under-stamped letter) would be available from memory. Similarly, Griggs and Cox (1982) attributed their "complete failure to replicate the good performance" (p. 413) with American subjects to the fact that the rule was not a part of their subjects' past experience. No such postal rule existed in the United States. Because the rule was familiar to our German subjects, availability theory therefore should predict a high percentage of *P & not-Q* responses.

Most important, availability theory should predict no difference in *P & not-Q* responses between perspectives. The rule is identical in both perspective versions, and the context stories are similar and of equal length; thus, falsifying instances should be equally available from memory. More generally, if frequency of co-occurrence of *P* and *not-Q* in the past, or its memory counterpart, is the criterion (see Pollard, 1982), then a perspective change should make no difference in *P & not-Q* responses. It is less clear what level of *not-P & Q* responses availability theory should predict (this would depend on how easily subjects can recall unsealed envelopes with a 1-mark stamp from memory).

Similarly, cheating and perspective are not concepts in PRS theory. Since the post office rule is an obligation rule, PRS theory should predict a high percentage of *P & not-Q* answers, and a very low percentage of *not-P & Q* answers, independently of the perspective.

& not-Q answers, and a very low percentage of *not-P & Q* answers, independently of the perspective.

Table 3 shows the results. The perspective change results in a strong decrease in *P & not-Q* responses, as predicted by SC theory, but not by PRS and availability theories. The decrease is about as large as that found for rules (1) and (2), namely 48%. But there were also 28% *not-P & Q* answers in the inverted perspective as opposed to 0% in the original perspective, a finding implied by none of the three theories. This percentage of *not-P & Q* responses is less than half of the amount found in the social contract rules with bilateral cheating options.

The second rule was similar to the "cholera" rule (Cheng et al., 1986):

(11) If a passenger is allowed to enter the country, then he or she must have had an inoculation against cholera.

One context story cued the subject into the perspective of an immigration officer at the international airport at Manila, whose task is to check whether any of four passengers has violated the rule. This context story was essentially the same as Cheng and Holyoak's "rationale version" of the cholera problem. We refer to it as the "original" perspective. The second context story cued the subject into the perspective of a passenger who makes a stop at Manila and would like to spend a few days there for pleasure. The passenger, who has no cholera inoculation, gets a hint from an informant that the first-aid physician at the airport ambulance station would be willing to perform and certify an inoculation quickly – of course, for money. The passenger is not sure whether such a certificate really will be accepted by the immigration officials, and suspects a trick to make money from tourists. Therefore the passenger checks information about four other passengers who were in the same situation. The subjects' task was to find out whether the rule was violated before. The four cards read: "is allowed to enter the country", "has no cholera inoculation", "is not allowed to enter the country", "has cholera inoculation".

In the cholera problem, the passengers can cheat (take the benefit of entering the Philippines without paying the costs of an inoculation: *P & not-Q*), and therefore SC theory predicts a high percentage of *P & not-Q* responses (and a very low percentage of *not-P & Q* responses) in the "original" perspective. The immigration office, however, cannot cheat. The *not-P & Q* combination means that a passenger who has a cholera inoculation is not allowed to enter. This behaviour does not imply "cheating" – although such behaviour could indicate unreliable or rude decisions, or simply the fact that there are other good reasons not to let this passenger enter the Philippines. Because of this unilateral cheating option, we get the same predictions as for the post office rule.

Table 3 shows that we also got the same results. The sharp drop in *P & not-Q* answers was again predicted by SC theory. The drop has now increased to 53

Table 3. *Reasoning about social contracts with unilateral cheating options* (n = *93)*

	Predictions			Results (in %)		
Perspective	Social contract	Availability	PRS	Post office	Cholera	Drinking
P & not-Q responses						
Original	High	High	High	80	85	92
Switched	Low	High	High	32	32	61
not-P & Q responses						
Original	Very low	Very low	Very low	0	0	0
Switched	Very low	Very low	Very low	28	23	11

Note: Numbers are percentages of subjects choosing *P & not-Q* or *not-P & Q* responses.

percentage points. And, again we find a similar difference in *not-P & Q* responses between perspectives. *Not-P & Q* means that the tourist paid the costs but did not receive the benefit. But paying the costs is not a benefit for the immigration office, as it would be in bilateral cheating situations.

The third rule we used was the following drinking age rule that holds in Germany:

(12) If a customer is drinking an alcoholic beverage, then he or she must be over 18 years old.

In one context story, the subject was cued into the perspective of a waitress in a pub, whose job is to check whether any of four young persons sitting at the next table has violated the rule. This is the law enforcement perspective used in earlier studies of the drinking age rule, and we refer to it as the "original" perspective. In this perspective, the rule elicited a *P & not-Q* response from about 75% of subjects tested (e.g., Griggs & Cox, 1982, 1983). In the switched perspective, subjects were cued into the role of a young customer who just became 18 years old. The customer knows that there are police controls and that the owner of the pub strictly follows the drinking law. In particular, when the owner is in a bad mood, he sometimes refuses alcohol to young people, if it is hard to decide whether they are over 18 or not. In both versions, the subjects were presented information about four young persons sitting at the next table. The subjects' task was to find out whether the rule was violated. The four cards read: "customer drinks alcohol", "is 16 years old", "customer drinks Coke", "is 18 years old".

In the drinking age problem, only the customer can cheat. Note that cheating means here to take a benefit (drink alcohol) but not meet a *requirement* (being 18 years old). Age is not a cost in the sense the term was used in the preceding rules. SC theory predicts a high percentage of *P & not-Q* responses (and a very low percentage of *not-P & Q* responses), if the original perspective is cued. The owner of the pub, however, cannot cheat. That a customer is over 18 years old is not a benefit for the pub owner. The owner may have good reasons for refusing alcohol to someone who is over 18 – for instance, the customer looks too young and has no identification to prove his or her age. But this is not cheating in the sense of taking a benefit without paying costs or satisfying a requirement. Therefore, in the switched perspective, SC theory should give the same predictions as the two preceding rules.

Table 3 shows that the results had the same tendency as in rules (10) and (11), but the differences between perspectives were only about half the size found in the two preceding rules.

How can we understand the modest proportions of *not-P & Q* responses in the switched perspective of the post office and cholera rules? And why do responses in the switched perspective of the drinking age problem differ? None of the theories above implies these results, and our suggestions are post hoc. One suggestion is the concept of free-rider parties. In the switched perspective of the cholera problem, *not-P & Q* does not mean that the immigration office cheats (see above), but it can mean that a third, free-rider party cheated the passenger: the first-aid physician and the informant (the person who suggests to the passenger that he or she go to the physician) may play a trick on passengers. They may take the benefit of his or her money while offering a certificate that does not give the passenger the benefit of entering the country. Thus, a third party who takes advantage of the social contract, not the actual partner in the social contract, could cheat. The switched perspective of our cholera problem allows for such an interpretation. Thus we cannot exclude the possibility that a cheater-detection algorithm, although directed at a free-rider party, accounts for the modest number of *not-P & Q* responses.

The post office and drinking problems, however, do not invite a similar interpretation. In the post office problem, one also could identify a third party: all the employees in the firm who post letters. (Recall that one employee, into whose perspective the subjects were cued, checks these letters.) There was, however, no indication in the context story that the employees could derive a benefit from overstamping unsealed envelopes (*not-P & Q*), nor is this an interpretation that would suggest itself to our subjects from their everyday experience. What might help to understand the modest number of *not-P & Q* responses in the post office problem is the notion that social contracts provide some *net* benefit for both parties, that is, positive profit margins (Cosmides & Tooby, 1989). Overpaying postage (*not-P & Q*) may threaten positive profit margins from the perspective of the sender. But this post hoc explanation is at best a direction to be developed in further research – for instance, within the context of Manktelow and Over's (1991, in press) notion of utilities in deontic reasoning.[7]

Response distribution

Residual responses were few in the original perspectives. In the drinking problem, all residual responses were *P* responses (9%); *P*, *Q*, & *not-Q* or "all four" were chosen in neither of the three problems. In the switched perspective, residual responses were similar in the post office and cholera problems, with "all

[7] One reviewer suggested that some subjects might perceive a second social contract in the postal situation: the boss is paying the employees to do their job carefully, and the attention they must pay to their task, for example stamping envelopes, is a cost for the employees and a benefit for the boss. Employees can cheat their boss by not paying attention with a risk of over- or understamping. Thus, subjects can perceive their task in the switched rule as one of checking that the boss is not being cheated, either by overstamping (*not-P & Q*) or understamping (*P & not-Q*). Thus, the four-card selection should be frequent – and it is the most frequent (17%) residual response. Two-card selections should be sensitive to the relative risk of over- or understamping: if 1-mark stamps are more ordinary, the probability of overstamping may be considered higher, and *not-P & Q* more likely to be relevant for detecting cheaters. In the present case there was no information given on relative risks of over- or understamping, and consistent with this the percentages of *P & not-Q* and *not-P & Q* were about the same.

four" being the most frequent (17% and 13%, respectively). For the drinking problem, the most frequent residual response was again *P* (13%).

Summary and discussion: Part II

We distinguished two kinds of social contracts: contracts with bilateral cheating options and ones with unilateral cheating options. Perspective change (from party B to party A) in social contract rules with bilateral cheating options led to a near-perfect reversal of *P & not-Q* responses into *not-P & Q* responses. Results followed the predictions of SC theory (in the present, revised version where the concepts of a social contract rule, a cheater-detection algorithm, and a person's perspective are separated). Manktelow and Over (1991, in press) reported similar effects of perspective change in permission rules. Reasoning about social contracts with unilateral cheating options followed the predictions of SC theory in the original perspective. In the switched perspective, however, results were only in part consistent with SC theory. Perspective change in social contracts with unilateral cheating options led to a half-way reversal, that is, almost the same number of *P & not-Q* and *not-P & Q* responses in the post office and cholera problems. Only in the cholera problem could we offer a post hoc explanation for the moderate number of *not-P & Q* answers. Results of switching perspectives contradict availability theory and cannot be accounted for by PRS theory.

Was there any evidence of mental logic in our experiments? In the post-experimental interview, six subjects said that all stories "had the same formal structure", and that they had always given the same, logically correct answer. We checked their actual answers, and these subjects had indeed always chosen the *P & not-Q* answer. No other subjects did. This group of subjects all came from mathematics and the natural sciences, except for one, who was a psychology student. When asked what role the content of the problems played in their reasoning, these subjects reported that they had experienced a dissonance between abstracting logical structure (using symbolic representation) and paying attention to the content: "Sometimes, in fact, a different answer would have been more meaningful", as one subject put it. Nevertheless, they asserted, they carefully followed the instructions and simply checked whether a rule was violated or not.

Thus a small group of students consistently solved the selection tasks with a mental logic, and without any deviation for a whole series of 12 problems. From the interviews, however, it is also clear that in some problems they experienced conflict between their mental logic and what their intuition otherwise would have directed. This experienced dissonance makes us hesitant to introduce the term "natural logic" here instead of "mental logic". The mental logic of these subjects followed propositional logic. These subjects' responses do not support SC theory, nor do they support PRS or availability theory.

Reasoning by whom and for what purpose: Perspective and cheating detection

Cheng and Holyoak (1989) and Pollard (1990) launched a sharp attack on Cosmides' conclusion that a "social contract will elicit a high percentage of the predicted social contract response, *P & not-Q*, but the non-SC permission rule will not" (1989, p. 243). What Cosmides called a "non-SC permission rule" is "a rule that lacks the cost–benefit structure of a social contract, but that does fit the action–precondition representation of a permission" (p. 243).

The debate focused on the question of whether a given rule is a social contract or a non-SC permission rule, or something in between. Cheng and Holyoak (1989) argued that Cosmides goes back and forth between what they call "actual social exchange" and "pseudo-exchange". Only an actual social exchange has a symmetrical cost-benefit structure, whereas in a pseudo-exchange costs are replaced by "requirements". For instance, in Cosmides' standard social contract version of the cassava rule (see rule 1), eating cassava root instead of molo nuts is a benefit, but having a tattoo (a sign of being married) is more appropriately called a requirement than a cost. In Cheng and Holyoak's view, pseudo-exchange rules such as the cassava rule blur the line Cosmides wants to draw between social exchange and permission rules, since pseudo-exchange rules are in between social exchange rules and permission rules. Since Cosmides' concept of a "requirement" seems to be almost identical to Cheng and Holyoak's "precondition" in permission schemas, the only remaining difference is that the "benefit" in pseudo-exchange is a subset of "action to be taken" in PRS theory. Thus, in the discussion to date, the critical theoretical question appeared to be whether a given rule is a social exchange rule, a permission rule without social exchange, or something in between.

We have argued, and experimentally demonstrated, that this is not the crucial issue. For instance, all the rules (3) to (6) were social contracts in Cosmides' terminology, and permissions or obligations in Cheng and Holyoak's terminology. But none of them elicited a high percentage of *P & not-Q* responses *unless* the subject was cued into the perspective of one participating party, and the other party had the option of cheating. *The crucial issue about social contracts is the cheating option, which in turn is a function of the perspective.* The fact that a rule is perceived as a social contract, either as a permission or an obligation rule, turned out *not to be a sufficient condition* for eliciting a high percentage of *P & not-Q* responses and the other striking results reported by Cosmides.

Having disentangled the concept of a social contract from that of a cheater-detection algorithm, and having made explicit the concept of perspectives in social contracts, we propose the following revised predictions of SC theory:

> If a person represents one party in a social contract, and the other party has a cheating option, then a cheater-detection algorithm is activated that searches for information of the kind "benefit taken & costs not paid (requirement not met) by the other party".

Applied to the Wason selection task, this predicts:

> If a selection task (rule and context information) (1) cues a person into the perspective of one party engaged in a social contract, and (2) the other party has a cheating option, then a cheater-detection algorithm is activated that selects the conjunction of cards that define being cheated, that is, "benefit taken and costs not paid (requirement not met) by the other party".

As follows from the revised prediction, the fact that a rule is perceived as a social contract (or as a permission) is *not sufficient* for a large percentage of *P & not-Q* responses. This contradicts the prediction of both Cosmides and of Cheng and Holyoak, who proposed that if a rule is perceived as a social contract or a permission/obligation, respectively, then this would be a sufficient condition. It is the pragmatics of who reasons from what perspective to what end (e.g., cheating detection) that seems to be sufficient. Although Cosmides' prediction is at odds with this result, and although the distinctions between perspectives, and between social contract rules and Darwinian algorithms were not part of Cosmides' experiments, they nonetheless underlie SC theory (e.g., Cosmides 1985, Ch. 5; Cosmides, 1989, p. 235; Cosmides & Tooby, 1989, p. 85).

Recently, other studies have provided independent evidence for the importance of several pragmatic aspects in the selection task. Manktelow and Over (1991, in press) investigated perspective change with two social contract rules and found effects of perspective similar to those in the present study. Light, Girotto, and Legrenzi (1990) showed that children's selections depend on the *kind* of cheating (e.g., selfish, nepotist) they expect.

Research on deductive and probabilistic reasoning: Some common issues

There has been little interaction between research on deductive reasoning – the heading under which the selection task is usually classified – and research on probabilistic reasoning. Yet there are striking parallels in the ways these two fields have developed. This parallel development was driven by the same two questions: "What does it mean to be rational?" and "How shall we understand the role of content and structure in reasoning?"

In the 1960s, both Peter Wason's (1966) selection task and Ward Edwards' (1968) probability revision task were introduced as analogies to scientific hypothesis testing. The selection task was seen as an analogy to Karl Popper's logic of falsification (see Johnson-Laird & Wason, 1970); the probability revision task was seen as an analogy to the Bayesian version of scientific hypothesis testing (see Edwards, Lindman, & Savage, 1963). In the former, subjects were asked to search for information that could falsify a conditional; in the latter, subjects were asked to revise the prior probability of a hypothesis into a posterior probability.

Based on a large body of research on both tasks, however, it was concluded that human reasoning did not generally follow these logics.[8]

If the mind is neither a Popperian nor a Bayesian, what then? How do people reason in the selection and probability revision tasks? In both research programs, the focus shifted from the logical structure to the content of the tasks:

> For some considerable time we cherished the illusion . . . that only the structural characteristics of the problem mattered. Only gradually did we realize first that there was no existing formal calculus which correctly modelled our subjects' inferences, and second that no purely formal calculus would succeed. Content is crucial . . . (Wason & Johnson-Laird, 1972, pp. 244–245)

In their research on probabilistic reasoning, Kahneman and Tversky (1982) reached the same conclusion: "Human reasoning cannot be adequately described in terms of content-independent formal rules" (p. 499).

Nevertheless, the content-independent formal rules were retained in both programs, not dropped. In research on the selection task, propositional logic was retained as the yardstick of correct reasoning, and researchers attempted to explain why certain contents "facilitate" logical reasoning. The "availability" hypothesis, for instance, assumed that deviations from logic become less frequent when the content of a conditional becomes more familiar. Similarly, in research on probability revision it was the *deviation* between subjects' judgments and the "correct" answer derived from Bayes' theorem that became the *explanandum* – not subjects' judgments in and of themselves. Depending on the direction of the deviation, the explanandum was called "conservatism" (Edwards, 1968) or "base rate fallacy" (Bar-Hillel, 1984), and seen as an error of reasoning. Much of the research of the 1980s was concerned with identifying features of the content – such as its vividness, specificity, and causality – that would "facilitate" Bayesian reasoning.

Thus, in both programs, similar answers were given to two key issues in human reasoning:

(1) What does it mean to be rational? The answers proposed were content-independent formal theories. For the most part, good reasoning was reduced to propositional logic and a simplistic version of Bayesianism (see Gigerenzer, 1991a, 1991b).

(2) How shall we understand the role of content and structure in reasoning? In the dominant program on probabilistic reasoning, the heuristics and biases program (e.g., Tversky & Kahneman, 1974), the formal structure of a

[8] These results were taken by many colleagues as a demonstration of reasoning errors in human subjects. However, we should remember that social scientists' reasoning does not generally follow Popperian or Bayesian logic, either. One has to go a long way to find a social scientist who uses Bayes's theorem for testing his or her scientific hypotheses or follows Popper's (1959) prescriptions, for example, to design general theories with very specific consequences and to attempt to falsify one's own theories by testing these consequences. Research on the Wason selection task has been no exception to this pattern.

> reasoning problem has defined correct reasoning, whereas the content has been used to explain reasoning errors. Until recently, most research on the selection task was based on a similar assumption.

Why has there been so little progress in understanding the reasoning processes elicited by both tasks, despite an avalanche of studies since the 1960s? We believe one major reason to be the answers given to these two questions. These answers presuppose a *simple* dichotomy between structure and content, which is subsequently used to separate good from bad reasoning. However, this presupposition holds neither for modern logics nor for modern probability theory. Modern logics offer many structures, including deontic, three-valued, poly-valued, and modal logics (see Blau, 1978; Lewis, 1974; Strawson, 1952; von Wright, 1986); so does modern probability theory, including various Bayesian, frequentist, and non-additive theories (see Gigerenzer et al., 1989; Hacking, 1965; Shafer & Pearl, 1990). Each of these theories divides the information presented in a reasoning problem into two parts: relevant and irrelevant information. Relevant information corresponds to what is seen as the problem's structure. Irrelevant information corresponds to unimportant content. Different theories, however, make different divisions.

For instance, from the point of view of propositional logic, the relevant structure of the phrase "if *P* then *Q*" is that of a material implication, but it appears irrelevant whether the implication is an indicative or a deontic conditional. Consequently, researchers who adopted propositional logic as the yardstick of good reasoning have tended to ignore the distinction between selection tasks that involve an indicative conditional (e.g., "If there is an 'A' on one side, there is a '3' on the other") and a deontic conditional (e.g., a social contract). In contrast, in deontic and other modal logics, this and other differences constitute crucial conceptual distinctions. Thus, what counts as relevant structure and as irrelevant content differs among theories of logic.

The same holds for theories of probability. For instance, the Bayesian point of view adopted by the heuristics-and-biases program does not distinguish between relative frequencies and single-event probabilities. Consequently, researchers have tended to ignore that distinction in probability revision tasks. In contrast, for the frequentist interpretation of probability, this distinction is fundamental; and so it seems to be for understanding intuitive reasoning as well (Gigerenzer, 1991b; Gigerenzer, Hoffrage & Kleinbölting, 1991).

The general point is that there is no simple and unique division line between structure and content, or between information relevant and irrelevant to rational reasoning. What counts as the relevant structure for reasoning about a domain therefore seems to need a domain-specific theory. Thus we need to define what the relevant structural properties are – modal operators, prior probabilities, likelihoods, perspectives, cheating detection, and the like – rather than to leave this job to one out of many possible logics, usually selected by convention.

We began this article by opposing content-independent formal theories of reasoning such as propositional logic with content-dependent theories such as Darwinian algorithms adapted to specific domains. We can now see that this opposition is not a dichotomy; there is a continuum between these poles. For instance, the domains of PRS theory – conditional permission, obligation, and causality – can also be dealt with in terms of modal operators, and such modal logics are available (e.g., Lewis, 1974; Rips, 1990). Thus, logics and probability theories can be domain specific too, although the contours of these domains are not drawn by evolutionary or ecological argument. This continuity suggests a potential for a fruitful exchange between logical structures on the one hand and domain-specific analysis, ecological or evolutionary, on the other – a potential much larger than a simple opposition between truth-table logic and Darwinian algorithms would suggest. For instance, proponents of a "natural logic" such as Martin Braine have recently moved to incorporate pragmatic principles – such as Grice's (1975) "cooperative principle" and Geis and Zwicky's (1971) "invited inferences" – into a theory of reasoning about conditionals (Braine & O'Brien, 1991). To date, however, these pragmatic sources of reasoning still function as an appendix to logical inference schemas.

If human reasoning is, to some important degree, an adaptation to specific environments (where environments include social environments), then ecological analyses of reasoning mechanisms as adaptations to structures of important present environments, and evolutionary analyses of reasoning mechanisms as adaptations to structures of important past environments, are indispensable. SC theory is one important step. However, in much contemporary research on probabilistic and deductive reasoning, the absence of an analysis of the structure of environments is still a defining feature (Gigerenzer & Murray, 1987, Ch. 5).

The dissociation of research on deductive and probabilistic reasoning is as obsolete as the parallels between the two programs are striking. We expect that the two fields will converge in the next few years, as a consequence of the growing role of pragmatic principles, such as perspectives, cost–benefit analyses, and cheating detection, in both fields.

References

Bar-Hillel, M. (1984). Representativeness and fallacies of probability judgment. *Acta Psychologica*, *55*, 91–107.

Blau, U. (1978). *Die dreiwertige Logik der Sprache*. Berlin: De Gruyter.

Boole, G. (1958). *An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities*. New York: Dover (original work published in 1854).

Braine, M.D.S., & O'Brien, D.P. (1991). A theory of *If*: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, *98*, 182–203.

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

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

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

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

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

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

Daston, L. (1988). *Classical probability in the Enlightenment.* Princeton, NJ: Princeton University Press.

Duyne, P.C. van (1974). Realism and linguistic complexity in reasoning. *British Journal of Psychology*, *65*, 59–67.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment.* New York: Wiley.

Edwards, W., Lindman, H., & Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*, 193–242.

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

Geis, M., & Zwicky, A.M. (1971). On invited inferences. *Linguistic Inquiry*, *2*, 561–566.

Gigerenzer, G. (1991a). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, *98*, 254–267.

Gigerenzer, G. (1991b). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology*, *2*, 83–115.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.

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

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life.* Cambridge, UK: Cambridge University Press.

Girotto, V., Blaye, A., & Farioli, F. (1989). A reason to reason: Pragmatic basis of children's search for counterexamples. *European Bulletin of Cognitive Psychology*, *9*, 297–321.

Girotto, V., Gilly, M., Blaye, A., & Light, P. (1989). Children's performance in the selection task: Plausibility and familiarity. *British Journal of Psychology*, *80*, 79–95.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics, III: Speech acts* (pp. 41–58). New York: Academic Press.

Griggs, R.A. (1984). Memory cueing and instructional effects on Wason's selection task. *Current Psychological Research and Reviews*, *3* (4), 3–10.

Griggs, R.A., & Cox, J.R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*, 407–420.

Griggs, R.A., & Cox, J.R. (1983). The effects of problem content and negation on Wason's selection task. *Quarterly Journal of Experimental Psychology*, *35A*, 519–533.

Hacking, I. (1965). *Logic of statistical inference.* Cambridge, UK: Cambridge University Press.

Inhelder, B., & Piaget, J. (1958). *Growth of logical thinking: From childhood to adolescence.* New York: Basic Books.

Johnson-Laird, P.N. (1982). Thinking as a skill. *Quarterly Journal of Experimental Psychology*, *34A*, 1–29.

Johnson-Laird, P.N., Legrenzi, P., & Sonino Legrenzi, M. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, *63*, 395–400.

Johnson-Laird, P.N., & Wason, P.C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, *1*, 134–148.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 493–508). Cambridge, UK: Cambridge University Press.

Laplace, P.S. (1951). *A philosophical essay on probabilities*. New York: Dover (original work published 1814).

Lewis, D. (1974). Semantic analyses for dyadic deontic logic. In S. Stenlund (Ed.), *Logical theory and semantic analysis* (pp. 1–14). Dordrecht: Reidel.

Light, P., Girotto, V., & Legrenzi, P. (1990). Children's reasoning on conditional promises and permissions. *Cognitive Development*, *5*, 369–383.

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

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

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

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

Manktelow, K.I., & Over, D.E. (in press). Obligation, permission, and mental models. In Y. Rogers, A. Rutherford, & P. Bibby (Eds.), *Models in the mind*. London: Academic Press.

Pollard, P. (1982). Human reasoning: Some possible effects of availability. *Cognition*, *12*, 65–96.

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

Popper, K.R. (1959). *The logic of scientific discovery*. New York: Basic Books (originally published in 1935).

Rips, L.J. (1990). Reasoning. *Annual Review of Psychology*, *41*, 321–353.

Shafer, G., & Pearl, J. (Eds.) (1990). *Readings in uncertain reasoning*. San Mateo, CA: Morgan Kaufmann.

Strawson, P.F. (1952). *Introduction to logical theory*. London: Methuen.

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

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

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

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

Wason, P.C. (1983). Realism and rationality in the selection task. In J.St.B.T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.

Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Wright, G.H. von (1986). Truth, negation, and contradiction. *Synthese*, *66*, 3–14.

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

Yachanin, S.A., & Tweney, R.D. (1982). The effect of thematic content on cognitive strategies in the four-card selection task. *Bulletin of the Psychonomic Society*, *19*, 87–90.