ELSEVIER

Cognition 57 (1995) 31–95

COGNITION

Relevance theory explains the selection task

Dan Sperber[a,*], Francesco Cara[a], Vittorio Girotto[b]

[a]*Centre de Recherche en Epistémologie Appliquée, Ecole Polytechnique, 1 rue Descartes, 75005 Paris, France*

[b]*CREPCO, Université de Provence, 29 Avenue Robert Schuman, 13100 Aix en Provence, France*

Received July 5, 1994, final version accepted February 3, 1995

Abstract

We propose a general and predictive explanation of the Wason Selection Task (where subjects are asked to select evidence for testing a conditional "rule"). Our explanation is based on a reanalysis of the task, and on Relevance Theory. We argue that subjects' selections in all true versions of the Selection Task result from the following procedure. Subjects infer from the rule directly testable consequences. They infer them in their order of accessibility, and stop when the resulting interpretation of the rule meets their expectations of relevance. Subjects then select the cards that may test the consequences they have inferred from the rule. Order of accessibility of consequences and expectations of relevance vary with rule and context, and so, therefore, does subjects' performance. By devising appropriate rule-context pairs, we predict that correct performance can be elicited in any conceptual domain. We corroborate this prediction with four experiments. We argue that past results properly reanalyzed confirm our account. We discuss the relevance of the Selection Task to the study of reasoning.

1. Introduction

In 1966, Peter Wason introduced what is now known as the Wason Selection Task (see Fig. 1). All true versions of the task share the same four-component structure:

(i) An introduction (sometimes in a narrative form).
(ii) A conditional statement, with the linguistic form "if P, then Q", known as the "rule".

* Corresponding author. E-mail: sperber@poly.polytechnique.fr

SSDI 0010-0277(95)00666-4

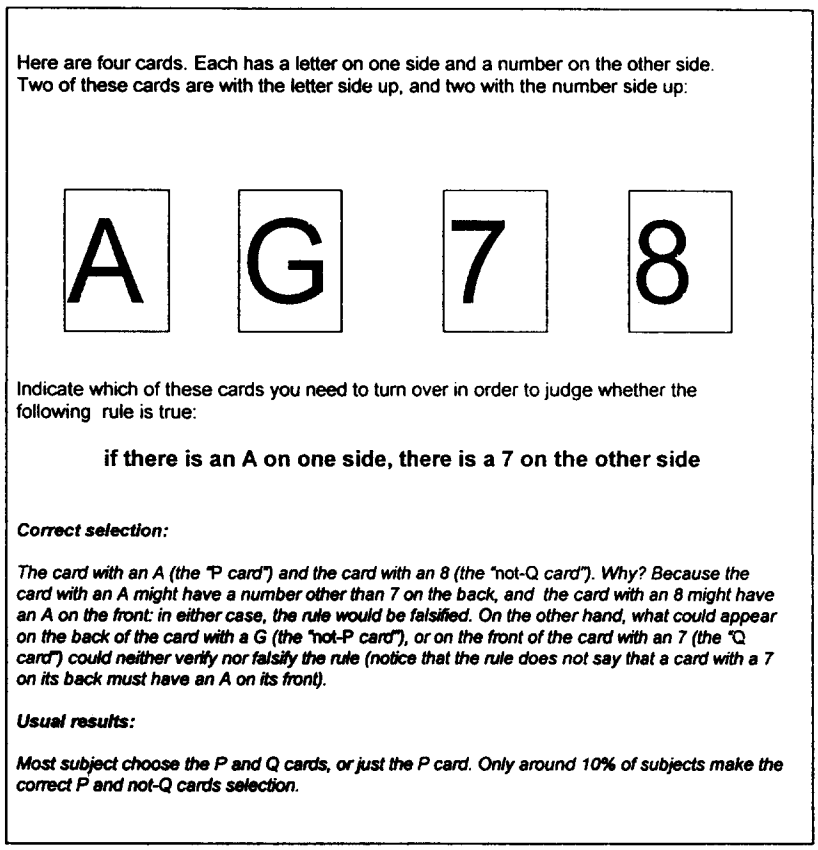

Fig. 1. A standard abstract version of the Selection Task.

(iii) Four cards (or, more commonly, a picture of four cards), each representing a case that either satisfies or does not satisfy P, and either satisfies or does not satisfy Q. The information regarding the satisfaction of Q is hidden from sight on two of the cards, and the information regarding the satisfaction of P is hidden from sight on the two other cards. The four cards are called the *P* card (representing a case where the antecedent of the rule is visibly satisfied), the *not-P* card (representing a case where the antecedent of the rule is visibly not satisfied), the *Q* card (representing a case where the consequent of the rule is visibly satisfied), and the *not-Q* card (representing a case where the consequent of the rule is visibly not satisfied).

(iv) The instruction to select all and only those cards where the hidden information must be made visible in order to judge whether the rule is true (or, in some versions, is being obeyed).

The Wason Selection Task has ben described in a recent textbook on human reasoning as "the most intensively researched single problem in the

history of the psychology of reasoning" (Evans, Newstead, & Byrne, 1993, p. 99). Is the Selection Task to the psychology of reasoning what the microscope has been to biology? Is its success due to the scientific advances it has made possible? Or should it be compared, rather, to the Rubik Cube, and its success accounted for by the fact that the psychology of this apparently simple task has proved particularly baffling, one explanation after the other being proved wrong?

At first, the challenge was just to explain why about 90% of subjects failed to make the correct selection of the *P* and *not-Q* cards. However, in the early 1970s, two new varieties of the task were reported that succeeded in eliciting a majority of correct selections. Jonathan Evans (1972) reported that a majority of subjects (about 60%) make the correct selection when the rule is of the form: "if P, then *not* Q", that is, when the consequent of the conditional is in the negative form. The other variety with which successful performance has been elicited has attracted much more attention. It involves not a descriptive, but a deontic conditional rule. Johnson-Laird, Legrenzi, and Legrenzi (1972) provided the first of a long series of examples of the fact that when the so-called rule is truly a rule in the ordinary deontic sense, that is, when the rule expresses a duty or a right resulting from social, contractual, or prudential arrangements, a majority of subjects correctly selects the *P* and the *not-Q* card (see Fig. 2 for a later version of this type of task).

Initially, however, Johnson-Laird et al.'s study was interpreted as evidence for what soon turned out to be too simple an explanation: the standard versions, it was first hypothesized, failed to elicit correct responses because of their abstractness, while versions with concrete, realistic content, close to people's experience, would "facilitate" reasoning and elicit correct

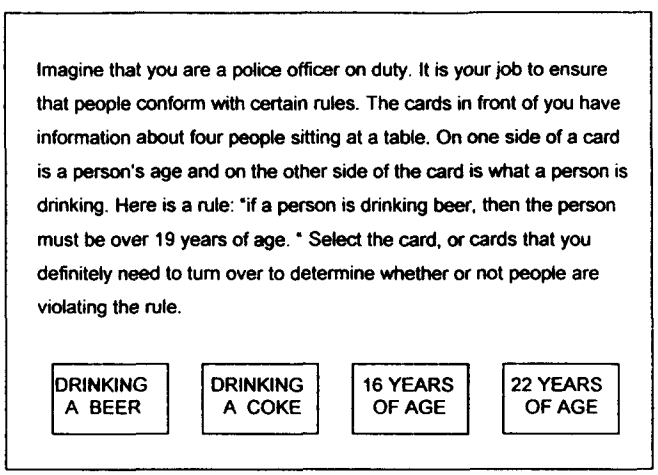

Fig. 2. A deontic version of the Selection Task: the drinking age problem (Griggs & Cox, 1982).

performance. This hypothesis proved wrong, however, when versions that were familiar and concrete but non-deontic failed to elicit the expected good performance (Manktelow & Evans, 1979) or failed to replicate (Manktelow & Evans, 1979; Wason & Shapiro, 1971), and when abstract deontic rules (Cheng & Holyoak, 1985, 1989) or unfamiliar ones (Cosmides, 1989; Girotto, Gilly, Blaye, & Light, 1989) were on the contrary successful.

It took some time too to realize that these deontic versions are logically different from the other, descriptive versions of the task (Manktelow & Over, 1990, 1991; Mosconi & D'Urso, 1974; Mosconi, 1990; Noveck & O'Brien, in press; Yachanin & Tweney, 1982). In the ordinary descriptive versions, the truth or falsity of the rule is what is in question, and what the cards selected are supposed to help establish. In the deontic versions, on the other hand, the truth of the rule (i.e., the fact that the rule is in force) is treated as axiomatic, and what subjects are expected to look for is not evidence of truth or falsity, but evidence of violation.

Given the logical difference between the ordinary, descriptive versions of the Selection Task, and the deontic versions, we concur with Griggs and Cox who argue that "to understand performance on the original Selection Task, it appears that we need to study it and not the numerous 'deontic' versions that have emerged during the past 27 years" (Griggs & Cox, 1993, p. 650). Our explanation of the task is built around the original descriptive version, and all the new experiments presented in this article likewise involve only descriptive tasks. We will return, however, to the deontic tasks in the concluding section, and argue that, in spite of their logical particularity, and in spite, also, of the domain-specific competences they may well evoke, they fall squarely under our general explanatory model.

There are competing theories to explain and predict performance for the negative-consequent-rule task (Evans, 1989; Johnson-Laird & Byrne, 1991; Oaksford & Chater, 1993), and for the deontic-rule task (Cheng & Holyoak, 1985; Cosmides, 1989; Johnson-Laird & Byrne, 1992; Gigerenzer & Hug, 1992; Manktelow & Over, 1991, 1992). There are also many insightful accounts, coming from different theoretical backgrounds (e.g., Evans, 1994; Fiedler & Hertel; 1994; Griggs & Cox, 1982; Kirby, 1994; Legrenzi, Girotto, & Johnson-Laird, 1993; Liberman & Klar, in press; Love & Kessler, in press;[1] Margolis, 1987; Oaksford & Chater, 1994; Platt & Griggs, 1993; Pollard & Evans, 1987; Shafir & Tversky, 1992; Wason & Green, 1984; Wason & Johnson-Laird, 1970) that throw light on the task in general, but mostly *ex post facto*. They may identify factors contributing to good performance, or to the selection of particular cards, but all these general accounts fall short of either predicting or ruling out good performance (i.e., more than 50% correct) on yet untested varieties of the task.

[1] There are encouraging convergences between our paper and those of Love and Kessler (in press) and of Liberman and Klar (in press), which we read only after having completed the present study.

What is most baffling today, then, is why good performance should be elicited by two unrelated varieties of the task (one of which, the negative-consequent-rule task, is often neglected in current discussions), and not by any other variety discovered so far.

In nearly thirty years of Selection Task research, there have been important advances in our understanding of the task, but it is open to question whether they have contributed much to our understanding of reasoning. In fact, a major step forward may well have been accomplished when Evans denied that reasoning was involved at all in subjects' usual performance (Evans, 1984, 1989). What subject do, according to Evans (1984, 1989, 1993), is merely to select those cards which appear *relevant* to them.

Evans mentions "an interesting degree of similarity" (Evans 1989, p. 26) between his use of the notion of relevance and that of "Relevance Theory" developed by Sperber and Wilson (1986), and, in recent papers (Evans, 1993, 1994), he uses the phrase "Relevance Theory" to name his current account of the Selection Task. However, Evans neither adopts Sperber and Wilson's notion of relevance, nor develops an explicit notion of his own. He considers rather individual factors of relevance, whether linguistic or contextual. Our approach may be seen as building on Evans' crucial insight that what the Selection Task does is elicit subjects' intuitive judgments of relevance, and as attempting to go further by availing itself of an explicit notion and theory of relevance. In the process, we are led to depart from some of Evans' more specific claims, and to exploit some other insightful contributions to the study of the task.

The word "reasoning" is often used by psychologists in a wide sense, as a synonym of "inference", the process of deriving conclusions from premises (i.e., in particular, of extracting more knowledge from knowledge already available). "Reasoning" can also be understood in a narrower sense (closer to ordinary usage), as referring to a deliberate and reflective inferential activity (to be distinguished from many other inferential processes that occur unconsciously, or even automatically). We agree with Evans that the performance of most subjects in the Selection Task is not the output of a conscious and deliberate activity of conditional reasoning. It is based, rather, on unreflective intuitions of relevance. We do believe, however, that these intuitions of relevance are the output of essentially unconscious *inferential* processes.

In this paper, we propose a general and predictive explanation of the Selection Task, a step-by-step solution to this psychologist's mind-twister. This explanation is based on a reanalysis of the task outlined in the second section, and on Relevance Theory (Sperber & Wilson, in press). The basic ideas and the relevance of Relevance Theory to the task are explained in the third section. According to the relevance-theoretic account of the Selection Task, most subjects' performance is wholly determined by expectations of relevance raised, in a predictable way, by the content and context of the

rule. Since relevance can be achieved in any conceptual domain, we predict, against majority opinion and against the apparent weight of previous evidence, that, by manipulating subjects' expectations of relevance, correct performance can be elicited *in any conceptual domain.*

Predicting that correct performance can be elicited in any domain does not commit us to the view that the cognitive processes that are at work in the Selection Task are domain-general. The processes involved are, we will argue, the standard processes of verbal comprehension. It is conceivable that these processes are, to some important extent, specific to comprehension. It is also conceivable that comprehension-specific mechanisms activate other domain-specific mechanisms when the information to be comprehended belongs to domains for which such mechanisms exist.

Four experiments, presented in the fourth section, serve to elaborate and confirm our predictions. The explanation is then shown in the fifth section, to explain past results on the Selection Task. In the Conclusion, we briefly discuss the relevance of the Selection Task to the study of reasoning.

2. Reanalyzing the Selection Task

2.1. The Selection Task is not a conditional reasoning task

The Selection Task has been from the start, and still is generally described as a conditional reasoning task (e.g., Wason, 1968; Evans, 1982; Johnson-Laird & Byrne, 1991; Rips, 1994). The poor performance generally observed is often considered as evidence of the absence of a basic conditional inference schema (viz., *modus tollens*) from the natural deductive repertoire of logically naive subjects (see, for instance, Anderson, 1986; Braine & Rumain, 1983; Smith, Langston, & Nisbett, 1992; Wason, 1968).

A standard conditional reasoning task is one where subjects are instructed to perform or evaluate a conditional syllogism, that is, a syllogism with a conditional statement of the form [if P, Q][2] as its major premise, and a statement of the form [P], [not-P],]Q], or [not-Q] as its minor premise. At first sight, it might seem that the Selection Task is a variation on that theme, with the rule providing the major premise and the visible side of the cards providing each a different minor premise. However, subjects' instructions are not to infer from the rule and the visible side of each card what must be on the other side of the card, nor to perform any other kind of conditional syllogism. Their instruction is to name all and only those cards that must be

[2] Here are the typographical conventions we have adopted: we use square brackets for logical forms, quotation marks for linguistic forms and statements, and parentheses to resolve structural ambiguities wherever necessary. We use roman "P" and "Q" for propositions and italic "*P*" and "*Q*" for properties or features.

turned over in order to ascertain whether the rule is true (or is being followed). Subjects are not instructed to find the solution by means of a deduction, let alone by means of a conditional deduction (such an instruction would be hard to convey effectively, anyhow).

Of course, the Selection Task has logically optimal, deductively valid solutions. The most obvious deductive solution does involve some conditional reasoning. It consists in trying to deduce from the rule and the visible side of each card what should be on the other side of the card. The two cards for which such a deduction is possible (i.e., the *P* card and the *not-Q* card) are, demonstrably, the two cards which should be turned over.

There are other deductively valid solutions. They may exploit, for instance, the logical equivalence between [if P, Q] and [(not-P) or Q] or that between [if P, Q] and [not-(P and (not-Q))]. Demonstrating these logical equivalences would involve conditional reasoning. However, subjects may have learnt these equivalences, or they may become aware of them under particular pragmatic conditions, without being able to demonstrate them. That is, they might arrive deductively at the correct solution, without engaging into any conditional reasoning. Solutions based on these logical equivalences have hardly received any consideration in the Selection Task literature.

We will argue that, when most subjects succeed in solving the task, their approach relies indirectly on the logical equivalence between [if P, Q] and [not-(P and (not-Q))] (or, more precisely, between the quantified counterparts of these logical forms). In so doing, we will be drawing on an insight of Johnson-Laird and Byrne (1991, p. 80) who have argued that subjects make the correct selection when they flesh out a model of the very combination of features *P-and-(not-Q)* ruled out by the conditional rule. We will add to this insight a general explanation of what may cause subjects to entertain the representation of the *P-and-(not-Q)* combination.

If we characterize the Selection Task by the type of instructions given to subjects, then it is not a deductive task. If we characterize it by the type of cognitive processes it evokes, then it is highly questionable that most subjects approach it as a deductive task (see below). Only if we choose to characterize the Selection Task by its logically optimal solutions, then might it be described as a deductive reasoning task, though, even then, not necessarily as a *conditional* reasoning task. But classifying tasks by their optimal logical solutions is of limited pertinence.[3] From a psychological point of view, what matters most is the cognitive abilities and the dispositions that a given task activates and tests.

[3] For instance, when Kahneman and Tversky (1972) asked subjects to estimate the product $1\times2\times3\times4\times5\times6\times7\times8$ or that of $8\times7\times6\times5\times4\times3\times2\times1$ under severe time constraints, this was, of course, not described as a calculus task, even though only calculus provides a failsafe algorithm to arrive at the solution.

2.2. The Selection Task is not an hypothesis evaluation task

Is the Selection Task a hypothesis evaluation task? At first blush one might say yes, and many people have. For instance, in his, 1982 textbook, Evans wrote that the Selection Task "is not *simply* a deductive problem. It may be regarded as an analogue of scientific hypothesis testing, in which subjects must appreciate the need to seek conditions which could *falsify* their hypothesis (cf. Popper, 1959)" (Evans, 1982, p. 157; see also, among many others, Johnson-Laird & Wason, 1970b; Klayman & Ha, 1987). The fact that some 90% of subjects fail to solve the task has been seen as evidence that most people are bad at hypothesis evaluation.

However, subjects are not asked to actually evaluate the rule, but only to select evidence for evaluating the rule. Their answers are not classified as more or less correct or incorrect according to whether the evidence they select would give them a greater or lesser chance of evaluating the rule accurately. Actually, with such a criterion, a simple statistical analysis shows that most answers would have to be classified as close to correct.

In real life situations, when a conditional hypothesis is being considered, cases that satisfy the antecedent are much more likely to provide adequate evidence for or against the hypothesis, than cases that fail to satisfy the consequent. This fact is related to Hempel's famous Raven Paradox (Hempel, 1965): if one wants to test the hypothesis that "if it is a raven, then it is black," one is much better off inspecting ravens and checking whether or not they are black, than inspecting non-black things and checking whether or not they are ravens. Let us, nevertheless, leave aside consideration of ecological validity and go for arbitrariness and randomness. Consider a Selection Task where the *P* card and the *not-Q* card have independent and equal probabilities of falsifying the rule. Even then, an evaluation of the rule based only on the *P* card (i.e., judging the rule false if the *P* card falsifies it, and true otherwise) has a probability of being accurate equal to at least .75. With more ecologically valid problems, the probability of such a truth-evaluation being accurate would always be greater than .75.[4] Thus, unless the falsifying cards are displayed neither randomly, nor in an ecologically valid way, selections that include the *P* card provide enough evidence for evaluating the rule accurately at least three times out of four, and though this is not good enough, it is far from abysmal.

Now, more than 90% of selections elicited in standard versions include the *P* card. Yet most of these selections, notwithstanding the fact that they

[4] More precisely: assume that, for each individual test, four cards are being drawn in such a manner that the probability that the *P* card is a falsifier of the rule and the probability that the *not-Q* card is a falsifier of the rule are independent from one another. Call p the probability that the *P* card is a falsifier, and q the probability that the *Q* card is a falsifier. Call r the probability that an evaluation of the rule based only on the *P* card is accurate (i.e., the probability that the rule is true if and only if the other side of the *P* card is a case of *Q*). Then $r = 1 - q + pq$. For $p = q = .5$, $r = .75$. For $p = q \neq .5$, $r > .75$. For $p > q$, $r > .75$.

would give subjects enough evidence for accurately evaluating the rule most of the time, are classified as wholly incorrect. The relatively rare *P*, *Q*, and *not-Q* cards selections, or all four cards selections, even though they provide all the evidence every needed for correct evaluation, have been described as exhibiting "partial insight" (Johnson-Laird & Wason, 1970a, 1970b), but are nevertheless classified as incorrect. This is no mistake or oversight on the part of researchers: the actual task is twice removed from that of evaluating the rule; it is to select *all* the cards that might serve as evidence for evaluation, and to select *only* those cards.

It is hard to think of any real life hypothesis evaluation task in which all the potential pieces of evidence must first be selected, and only then examined, and where, moreover, having selected an irrelevant piece of information is failing at the task. Typically, hypothesis evaluation is sequential: it involves looking at some evidence and then deciding what further evidence to seek, if any. In most conceivable cases where all the evidence might have to be gathered before being properly examined (e.g., special archeological or forensic situations, or downloading from a data base), picking pieces of evidence which, on second thought, could not have been relevant anyhow, does no harm to hypothesis evaluation, and is much better than failing to pick actually relevant evidence.

In most subjects' cognitive experience, hypothesis evaluation is a daily occurrence. The Selection Task, however, will be their first encounter with a case where failing to select at once all and only the potentially relevant evidence counts as altogether failing. So, not only subjects' task is not to evaluate the rule, it is not even to say how they would go about evaluating some such rule, if they really had to.

2.3. The Selection Task is just a task of selection

If the Selection Task is not a conditional reasoning task, if it is not an hypothesis evaluation task, then what is it? The answer is, in a way, quite obvious. It is spelled out in the task's very name: the Selection Task is a task of selection. What has been generally missed, in spite of Evans (1984, 1989), is that the task is nothing but that. All that subjects are asked to do is select potentially relevant evidence, and that is just what they do.

Relevance Theory (Sperber & Wilson, 1986) focuses on the psychological processes that guide the selection of information relevant to inferential processes. If correct, it should be particularly well suited to help explain the task. If incorrect, it is particularly at risk of being disconfirmed by means of the task.

Before developing a relevance-theoretic explanation of subjects' performance, we must reanalyze some key features of the task. We will do so on simple and general logical and linguistic grounds, independent therefore from our particular theoretical perspective.

2.4. *The rule is a universally quantified conditional statement*

In the Selection Task literature, it is generally assumed without discussion that "if", as it occurs in the rule, is properly interpreted as material implication. This is controversial, and not irrelevant to understanding several aspects of subjects' performance. For reasons of space, we will not discuss this point here. More central to a proper understanding of the task is the known but neglected fact that the rule is a general, and not a particular, conditional statement.

The rule is standardly represented as something like (1):

(1) $[P \rightarrow Q]$ ("if P, Q")

Logical form (1) is an appropriate representation of *particular* conditional statements such as: "If *this* card [pointing at a particular card] has a vowel on one side, it has an even number on the other side." Form (1) is not, on the other hand, an appropriate representation of a Selection Task rule such as "if *a* card has a vowel on one side, it has an even number on the other side", which contains an implicit universal quantifier. The conditional rule is a *general* conditional statement better represented as (2):

(2) $[\forall x\ (Px \rightarrow Qx]$ ("for any item x, if x has the feature *P*, x has the feature *Q*")

where "x" ranges over the cards, or over the cases represented by the cards, and where "*P*" and "*Q*" represent features, or properties, such as *having a vowel on one side* or *drinking beer*. The fact that the conditional rule of the Selection Task expresses a general conditional proposition is obvious and has never been denied. However the implications of this fact have hardly ever been considered (but see O'Brien, 1994).

The distinction between particular conditionals of form (1) and general conditionals of form (2) matters to the analysis of the Selection Task for two reasons. First, the two types of conditionals raise different problems of interpretation. In the case of general conditionals, for instance, the range of the universal quantifier must be determined. Second, the truth-value of the two types of conditionals is assessable in different ways.

Many, possibly most versions of the Selection Task are ambiguous in the following manner: the rule may be understood as being about some definite or indefinite wide range of cards (or of cases), only four of which are being displayed (or represented), or just about the four cards displayed (or about the four cases they represent). In other terms the range of the implicit universal quantifier is often itself left implicit and can be understood in two ways. These two possible interpretations have quite different epistemological consequences.

When understood as being only about the four cards, the rule (interpreted

as material implication) is both falsifiable and verifiable, and the same *P* and *not-Q* cards that may singly falsify it may also jointly verify it. The *Q* card, on the other hand, is not just unable to falsify the rule, it is also irrelevant to its verification (nor is there any clear sense in which it could be seen as capable of inductively "confirming" the rule, when the rule is so easily verifiable).

When understood as being about a wide range of which the four cards are only a sample, the rule is falsifiable (by selecting the *P* and *not-Q* cards), but not verifiable. However, falsification and verification are not the only two possible epistemic goals. Non-demonstrative confirmation in the absence of demonstrative verification, or disconfirmation in the absence of falsification, are reasonable goals too. Inductive strategies aimed at probabilistic confirmation or disconfirmation of the rule may be quite rational and they may produce different results depending on information about the distribution of the *P*, *not-P*, *Q* and *not-Q* features in the range (see Klayman and Ha, 1987; Kirby, 1994; Oaksford and Chater, 1994).

Experimenters who have bothered to disambiguate the task have, most of the time, done so in favor of the narrow range (e.g. Cosmides, 1989; D'Andrade, 1989; Hoch and Tschirgi, 1985; Wason, 1966, 1977). Kirby (1994, Experiments 2 and 3) is a rare case of disambiguation in favor of the wide range. In our own experiments, we have systematically disambiguated the task in favor of the usual narrow range.

Are subjects sensitive to the distinction between the narrow, four-cards range, and the wide range interpretation of the quantifier? Researchers who see the task as one of hypothesis testing should say yes, or be prepared to impute yet more irrationality to their subjects. We, on the other hand, hold that the task is reasonably understood by subjects as one more rudimentary than hypothesis testing, as so rudimentary, in fact, that the quantifier's exact range does not really matter. We predict therefore that, in most cases, disambiguating one way, or the other way, or leaving ambiguous an otherwise identical version of the task would have no major effect on subjects' performance. This however would deserve some systematic study.

2.5. The instructions

For all versions of the task to be truly comparable, the instructions should always be the same, at least in substance. They should limit themselves to something like (3):

(3) "Indicate which of the following four cards you definitely need to turn over in order to judge whether the rule is true or false"

However, as we already pointed out, the content of the instructions in deontic versions of the task has to be different from (3) and may, at best, be something like in (4):

(4) "Indicate which of the following four cards you definitely need to turn over in order to judge whether the rule is obeyed or violated"

There are other, interesting but unnecessary, differences found in the formulation of the instructions, both in the descriptive and in the deontic versions. In particular subjects are sometimes asked to select the cards necessary to judge whether the rule is true (or is being obeyed); at other times they are asked to select the cards necessary to judge whether the rule is false (or is being violated). In yet other versions they are asked to select the cards necessary to judge whether the rule is true or false (or is being obeyed or violated). This particular variation has been studied (Griggs, 1984; Yachanin & Tweney, 1982), and found to have less than systematic effects. Instructions to look for violation seem to interact positively with other factors favoring correct response, but to have no significant effect on their own (Chrostowski & Griggs, 1985; Yachanin, 1986).[5]

In many deontic versions, the question asked is not (or not just) whether the rule is followed (or violated), but quite explicitly, whether there are cases of cheating represented on the cards. Here are examples of such instructions:

(6) "Did Big Kiku get away with cheating any of these four men? Indicate only those card(s) you definitely need to turn over to see if Big Kiku has broken his word to any of these four men" (Cosmides, 1989, p. 265).
(7) "Which [cards] do you need to check on – to turn over – to see whether anybody cheated?" (Light, Girotto, & Legrenzi, 1990, p. 374).
(8) "Which of the receipts must you turn over to see whether this store has not given customers what they are entitled to?" (Manktelow & Over, 1991, p. 99).

Subjects' responses to such instructions are an interesting study in their own right. They are not easily compared, however, with versions of the task with more standard instructions.

Still, it can be argued that these explicit instructions to look for cases of cheating are quite revealing: they achieve by the "brute force" of their explicitness what other deontic versions (some of which have been successfully used by the same authors we have quoted) achieve by relying on subjects' tendency to read into more standard instructions an *implicit* invitation to look for cases of cheating. We surmise (comforted by two pilot studies of our own) that the explicit look-for-cheatings instructions could be replaced in otherwise unchanged versions by more standard instructions without significantly worsening subjects' performance. If so, then the

[5] The remarkable experiments suggested by Margolis (1987) and studied by Griggs (1989), and Griggs and Jackson (1990), involve not only non-standard instructions, but also non-standard rules. They deserve a detailed discussion for which we have no space.

explicit content of the brute force versions might be seen as revealing the implicit content of the more standard ones.

From the psychologist's point of view, however, the difference between explicit instructions and implicit suggestions is crucial (even if explicitness is a matter of degree; see Sperber & Wilson 1986, chap 4). Typically, experimenters strive to formulate their instructions in such a manner that comprehension of their explicit content is unproblematic, and can be taken for granted. In effect, the explicit content of the instructions (rather than their physical properties) is treated as the experimental stimulus. On the other hand, when subjects interpret the instructions as conveying implicit suggestions, this is an aspect of their *response* to the stimulus.

Concerning the Selection Task generally, our argument will be that subjects succeed in solving it when they interpret the instructions as inviting them implicitly to check whether there are *P-and-(not-Q)* cards. Instructing them explicitly to do so should achieve the correct selection. Indeed, Legrenzi (1970) found that almost 80% of his subjects were able to solve the problem, when presented with the rule:

(9) It is not possible for there to be a vowel on one side of the card and an odd number on the other side.

In an interesting set of recent experiments, Platt and Griggs (1993) found similar results. This, however, is only indirectly relevant to explaining performance with the standard task, where the rule is in conditional form, and without helpful paraphrase.

We want to test the hypothesis that differences in subjects' performance with various versions of the standard task reflect differences in the *implicit* import they attribute to the rule. With this hypothesis in mind, it has been our rule, in our own experiments, to try and trigger specific interpretive processes in our subjects without ever making the expected interpretation of the instructions anywhere explicit, since this would have defeated our purpose.

2.6. The logic and pragmatics of the task

The differences in performance between various versions of the Selection Task have been attributed to "content effects" (Cheng & Holyoak, 1985; Cosmides, 1989; Griggs & Cox, 1982; Manktelow & Evans, 1979). The basic idea is that all versions of the Selection Task present the same logical problem with one and the same logical solution. If subjects were just guided by logical considerations, they should, then, give the same answer in all versions. Differences in performance between versions are therefore explained by the cognitive effects of logically irrelevant content elements found in the introduction, the rule, the cards or the instructions of the task. This "non-logical content" is seen as inhibiting correct performance in some

cases, and as facilitating it in others. The existence of these content effects has been seen as casting doubts on human rationality (e.g., Stich, 1990).

We want to argue, however, that the notion of "content effects" rests on too rigid a dichotomy between those aspects of the content that subjects should pay attention to, and those they should ignore. From a normative point of view, we are, or should be, interested in the soundness of the subjects' overall cognitive performance – in other words in their rationality – and not just in the mere logicality of their response. We do not mean to deny that deductive logicality is a component of rationality (in contrast with Evans', 1993, notion of "rationality2"; see also Evans, Over, & Manktelow, 1993). Rationality as generally understood involves logicality but has other components too, for instance the ability to allocate one's cognitive resources efficiently, or that of performing non-demonstrative inferences.

In the Selection Task, the overall cognitive performance includes the comprehension of the task. Comprehension is, to a large extent, an inferential process of a non-demonstrative kind (studied by pragmatics: see Bach & Harnish, 1979; Blakemore, 1992; Leech, 1983; Levinson, 1983; Sperber & Wilson, 1986; Sweetser, 1990). To the extent that it is inferential, it can be achieved with lesser or greater rationality. It is rational, in comprehension, to take into account all the information given. For instance, when subjects are provided detailed information about the frequency of the different types of cards, as in Pollard and Evans (1983) or Kirby (1994), it is rational for them to assume that this information is relevant to their performance of the task. It is rational to take into account not just the truth-conditional content of a text, but also the way this content is formulated. To be influenced, for instance, by the fact that the instructions talk of ascertaining falsity rather than of ascertaining truth may be a symptom of pragmatic soundness and subtlety, and not of irrationality.

Generally speaking, the dichotomy between logically relevant and logically irrelevant content is always relative to an interpretation of the task. Past discussions of subjects' performance have tended to focus on the task as already interpreted (by the experimenter). Only with such a narrow focus does the interference of the "non-logical" content appear radically unsound. But interpreting the task is part and parcel of performing it, and obeys criteria of rationality in its own right. The study of "content effects" is the study of sound cognitive processes that are by no means out of place in subjects' performance.

In fact, we will argue, pragmatic processes of comprehension automatically involve determining where relevance lies, and they serve us rather well in this respect in ordinary life. When subject fail at the task, it is because of over-confidence in these pragmatic processes, and in the intuition of relevance that they determine. When most subjects succeed at a particular version of the task, it is because the pragmatics of that version are such as to elicit intuitions of relevance that happen to yield logically correct selections.

2.7. Direct and indirect truth evaluation

In the standard Selection Task, subjects don't have to actually evaluate the truth of the rule, but they have to consider what kind of evidence would be needed to do so. Under certain conditions, observations, for example the observation of both sides of cards, can verify or can falsify a statement. From a psychological point of view, it is important that we distinguish *direct* and *indirect* verification and falsification because they involve different cognitive processes. With some statements, one can directly see, or otherwise *observe*, that they are true or that they are false. For other statements, the best one can do is *infer* from what one observes that they are true or that they are false.

How much can be observationally ascertained is a contentious matter, much discussed in philosophy of science (e.g., Hempel, 1965; Popper, 1959; Quine, 1960; Van Fraasen, 1980). A very conservative approach to the problem would consist in claiming that the content of a genuine observation can be an atomic proposition attributing perceptible feature to a perceptible individual item. On such a view, (10) could be the content of an observation, but not (11)–(14), which can only be arrived at inferentially:

(10) the patch of grass on which you are standing now is green
(11) the patch of grass on which you are standing now is not red
(12) the patch of grass on which you are standing now is green and wet
(13) the patch of grass on which you are standing now is green or red
(14) grass is green

On a more liberal view of observability – the one we will adopt here – (11) and (12) could also be the content of observations. The case of (13) is much more dubious. As for (14), it is the prototype of a proposition that can never be the content of an observation.

Even on a very liberal construal of observability, conditional statements, whether particular or general, are never directly verifiable of falsifiable.[6] One cannot observe that [if P, Q], or that [for any item x, if x has the feature *P*, x has the feature *Q*], nor can one observe their negations. Conditional statements are only indirectly truth-evaluable. Their verification or falsification necessarily involves inferential steps.

[6] Mental models theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) may seem to afford an even more liberal view of observability. In this theory, particular conditionals are represented by means of three mental models which are each equivalent to a conjunction and are therefore each directly verifiable. However, converting a conditional statement into appropriate models is, we would argue, an inferential process, and therefore the converted statement itself is only indirectly verifiable. Our view could easily be rephrased in model-theoretic terms: the truth or falsity of a statement is directly observable only if that statement requires just one model.

A *particular* conditional statement (understood as a material implication) can be *verified* by deriving it from an observationally evaluable statement. For instance if [if P, Q] is true, then it is logically implied by [not-P]. If [not-P] is directly verifiable, then [if P, Q] is indirectly verifiable.

Whether a *general* conditional statement (understood as material implication) is verifiable depends on whether it is possible to observe all the items in the range of the quantifier. When the quantifier of a general conditional ranges over an actually or practically infinite set of items, verification is altogether impossible (as illustrated by (14), the logical form of which is: [for all x, if x is an instance of grass, then x is green]). In the Selection Task, however, with the rule understood as ranging just over the four cards, verification is trivially possible. The same two cards which may singly falsify the rule, may also jointly verify the rule.

A conditional statement, whether particular or general, can be *falsified* either by deriving from it some observationally testable implication and observing that this implication is false, or, equivalently, by identifying a proposition that contradicts it and observing that this proposition is true. As an example of the first method, take together a particular conditional statement of form [if P, Q] to be evaluated, and a statement [P] already known to be true. Together [if P, Q] and [P] imply [Q]. If [Q] were observed to be false, then [if P, Q] would be indirectly falsified. As an example of the second method, take a particular conditional statement of form [if P, Q]. It contradicts [P and (not-Q)]. If [P and (not-Q)] were observed to be true, then [if P, Q] would be indirectly falsified. These two methods of falsification are logically but not computationally equivalent.

2.8. Spontaneous inference and reflective meta-inference

Are formal methods of truth evaluation available to ordinary subjects? The question can be understood at two levels. If the issue is whether ordinary subjects know the difference between demonstrative and non-demonstrative truth evaluation, and can recognize adequate means of verification and of falsification, then, surely, the answer must be no. We don't want to attribute to ordinary subjects a mastery of conceptual distinctions which emerged only progressively in the history of logic, and which can only be acquired through some serious and deliberate learning. If, on the other hand, the question is whether ordinary subjects can perform demonstrative truth evaluation through an intuitive awareness of relationship of implication and contradiction, then the answer we favor is yes. This answer is supported, in particular, by the present study of the Selection Task.

The meta-inferential ability to reflect on ways of evaluating the truth of a statement on the one hand, and the mere inferential ability to perform some truth evaluations on the other hand are two types of resources that can be put to work on Selection Task problems, with different results. The meta-

inferential approach is the only fail-safe approach to the Selection Task, the only approach that should give the correct solution in each and every case. We assume that the minority of subjects who consistently give the correct answer in all versions of the task use this meta-inferential approach. As one might expect, the size of this minority is affected by the level of formal-logical competence of the population tested (Gigerenzer & Hug, 1992; Hoch & Tschirgi, 1985; Jackson & Griggs, 1988; Lehman & Nisbett, 1990). Some experiments have aimed at increasing the size of this minority by encouraging a more reflective stance (Cheng, Holyoak, Nisbett, & Oliver 1986; Hoch & Tschirgi, 1985; Johnson-Laird & Wason, 1970a; Platt & Griggs, 1993; Wason, 1969). It is conceivable that manipulations yet to be discovered might transform what is still a minority (on strictly standard tasks) into a majority, and induce more subjects to adopt a systematic meta-inferential stance from which they would consistently solve the task.

Regarding the current majority of subjects, who give a correct answer only for some versions of the task, two types of explanations have been envisaged. The first explanation, initially favored by Johnson-Laird and Wason (1970b), is that certain versions of the task make it more likely that subjects will adopt the meta-inferential approach and succeed. In other terms, certain versions were thought to "facilitate" good reasoning of a very general kind. This notion of "facilitation" has been criticized (Manktelow & Over, 1990, p. 110) and generally abandoned. The other explanation, in general favor today, is that thematic versions of the task which elicit good performance evoke domain-specific mechanisms (acquired "pragmatic schemas" as suggested by Cheng & Holyoak, 1985, or innate "Darwinian algorithms" as suggested by Cosmides, 1989).

We will defend another, third explanation of the fact that the majority of subjects give a correct answer in only some versions of the task. We attribute the performance of these subjects to spontaneous inferential abilities involved in comprehension, and differentially activated in different contexts by considerations of relevance.

In brief, we assume that there are two main ways[7] in which subjects approach the task. A minority of subjects adopt a reflective meta-inferential approach. A majority of subjects adopt a relevance guided, spontaneous, inferential approach. From now on, we focus exclusively on this second approach. To do so, we must now briefly introduce Relevance Theory.

[7] There are, no doubt, still other ways in which subjects may approach the task. Generally speaking, the variety of responses suggests a variety of individual approaches. Thus some subjects may be trying to reconstruct the intentions of the experimenters, others may see this as a mere guessing task, etc. Note, however, that the type of approach we attribute to the majority of subjects may by itself account for at least part of the variety of responses. Relevance considerations are relative not just to the stimulus, but also to the subject's abilities, interest, and level of attention. The same task may therefore seem relevant in different ways to different subjects, and yield different responses, all nevertheless based on the same general approach to the task.

The relevance of an information to an individual:

- The greater the cognitive EFFECT resulting from processing the information, the greater its relevance
- The greater the processing EFFORT required for processing an information, the lesser its relevance

The First (Cognitive) Principle of Relevance:

Human cognitive processes are aimed at processing the most relevant information available in the most relevant way

The Second (Communicative) Principle of Relevance:

Every utterance conveys a presumption of its own relevance

Fig. 3. The basic tenets of Relevance Theory.

3. Relevance Theory and the Selection Task

3.1. The basic tenets of the theory (see Fig. 3)

In *Relevance: Communication and Cognition* (1986), Sperber and Wilson characterize the relevance of an information to an individual in the following manner. Any new information is processed by an individual in a context of already available beliefs and conjectures. If bring together this new information and this context yields cognitive effects that could not have been derived from the new information alone, or from the context alone, then this information is relevant in this context, and to the individual who brought this context to bear on this information. The cognitive effects that make an information relevant can consist, in particular, in the addition of new beliefs implied by the information given the context, or in the abandonment of old beliefs, contradicted by the new information given the context.

Relevance is a matter of degree. *Ceteris paribus*, the greater the cognitive effects resulting from the processing of an information by an individual, the greater the relevance of that information to that individual on that occasion. Achieving cognitive effects involves a cost in the form of processing effort. Processing effort affects negatively the degree of relevance. *Ceteris paribus*,

the greater the effort involved in achieving the cognitive effects that make some information relevant, the lesser its relevance.

To illustrate: suppose that Peter wants to go as soon as possible to Manchester by train and doesn't know the train schedule. He asks Mary, and she tells him one of three things:

(15) The next train to Manchester is at 5:30 p.m.
(16) The next train to Manchester is sometime after 4:00 p.m.
(17) The next train to Manchester is scheduled to leave 7500 seconds after 3:25 p.m.

The information provided in (15) is the most relevant: from it, Peter can infer that the earliest time at which he can travel to Manchester is 5:30, and whatever else follows for him from that first inference. Statement (16) is less relevant than (15), since it yields fewer inferences (the proposition that the train leaves at 5:30 p.m. implies the proposition that it leaves sometime after 4:00 p.m., but not conversely). Statement (17) also is less relevant than (15), but for effort reasons this time. Since 7500 seconds after 3:25 is 5:30, Peter may derive from (17) all and only the inferences that he could derive from (15). However, because of the greater processing effort required, this convoluted answer is less relevant than the straightforward, ordinary version.

Sperber and Wilson argue that human cognition is guided by consideration of relevance in the following way. People tend to pay attention, at any given time, to the most relevant information available to them at the time, and to bring to bear on it a context of assumptions that will maximize this relevance. This is singled out as "the First (or Cognitive) Principle of Relevance" in the postface of the 2nd edition of *Relevance* (Sperber & Wilson in press).

The drive toward the maximization of relevance is guided by expectations of effect and considerations of effort. When cognitive effects are expected from specific information, attention is directed to those stimuli in the environment from which the desired information is expected, whether or not these stimuli are in and of themselves salient. Similarly, retrieval mechanisms may search for not so easily accessible background knowledge, which may allow the derivation of the expected effects. However, the expected effects have to be important to offset the added effort of perception, of representation, and of retrieval involved in such cases. In the absence of specific expectations of effect, or even in the presence of weak expectations of effect, considerations of effort may play the central role in directing attention and the retrieval of background knowledge. Since the perception and conceptual representation of information involves a processing cost, the most salient and easily representable information at a given time is, *ceteris paribus*, likely to be the most relevant information at that time. Similarly, since accessing a context involves a processing cost, the most accessible

contexts at the time are, *ceteris paribus*, those in which relevance is likely to be maximized.[8]

In general, information available in the environment does not, of itself, raise an expectation of cognitive effects. If a piece of information happens to be salient and easily representable, then it will be attended to in the hope of cognitive effects that may justify the relatively small effort expanded. If it fails to yield such effects, it will quickly be dismissed from attention.

Communicated information differs from ordinary environmental information in that it standardly raises a definite expectation of relevance. A communicator implicitly requests the attention of her audience, for without such attention her communication could never succeed. Since attention goes and stays only where relevance is expected, the communicator manifestly wants her audience to presume that the information she is attempting to communicate is relevant enough to be worth the attention implicitly requested. In other terms, every act of communication conveys a presumption of its own relevance. This is what Sperber and Wilson called "the Principle of Relevance" in the 1st edition of *Relevance* (1986), and what in the 2nd edition (Sperber & Wilson, in press) they call "the Second (or Communicative) Principle of Relevance".

The addressee of an act of communication, and in particular of an utterance, might be skeptical regarding the presumption of relevance, and indeed, the information conveyed may turn out not to be relevant to him after all. Relevance Theory does not claim that communicators always try to be relevant to their hearer, let alone succeed, nor that addressees always trust the communicator to be relevant to them. The crucial claim is this: whether or not the presumption of relevance is warranted, whether or not it is accepted, the very fact that it accompanies an utterance helps determine the utterance's intended interpretation. The intended interpretation has to be such that the speaker could think that it would satisfy the expectation of relevance that she herself encouraged in the hearer by means of her utterance.

In fact, the Communicative Principle of Relevance warrants a definite comprehension strategy which, we claim, is spontaneously followed by any addressee of an utterance. Comprehension involves disambiguating, assigning reference, narrowing down or loosening literal meaning, establishing

[8] This description raises an obvious question. How might a cognitive system so influenced by considerations of least effort have a good chance of latching on the potentially most relevant information in its environment and of processing it in an optimal context? The best answer is that we are dealing with an evolved and adapted mind/brain with fine-grained domain-specific competences (see Tooby & Cosmides, 1992; Sperber, 1994). For such a system, salient information, both in perception and in memory, tends to be information the processing of which is most likely to bring about significant cognitive effects. This is not to say, however, that attention and memory mechanisms actually succeed in maximizing relevance: the environment – and in particular a rapidly changing modern cultural environment – is likely to present many challenges with which these slowly evolved mechanisms are unequipped to cope.

illocutionary force, and identifying the cognitive effects which were intended by the speaker to make the utterance appear relevant to the hearer. Only this last aspect – the identification of intended cognitive effects – will be considered in the present study, but all aspects of comprehension are guided in the same manner by considerations of relevance.

The Communicative Principle of Relevance states, we said, that any utterance conveys a presumption of its own relevance. This presumption of relevance has an effect side and an effort side. On the effect side, the presumption is one of adequate effects: the effects must be sufficient to justify the effort required by the processing of the utterance. Note that the presumption is not one of maximal effects: the speaker has her own motivation in speaking and she will convey what she is most interested in conveying, which is not necessarily what the hearer would be most interested in hearing. On the effort side, it is both the speaker's and the hearer's interest that comprehension should require as little effort as possible. This is so for two reasons: first, the lesser the effort, the greater the relevance and therefore the greater the willingness of the hearer to pay the necessary attention; and, second, less effort means greater chance of successful processing. On the effort side, then, the presumption is one of minimal effort (for more detail see Sperber & Wilson, in press).

Given the presumption of relevance so understood, a rational comprehension strategy consists in:

(i) considering possible cognitive effects in their order of accessibility (i.e., following a path of least effort); and
(ii) Stopping when the expected level of relevance is achieved (or appears unachievable).[9]

Following a least-effort path in accessing possible effects is rational given the presumption that the speaker has tried to minimize the effort needed in order to arrive at the intended interpretation. Stopping when the expected level of effect is achieved is rational given that, on the effect side, the presumption is one of merely adequate effects.

Here is an informal example of a comprehension process stopping when the expected level of relevance is achieved:

(18) *Peter:* Do you want to go the party at the Smiths?
Mary: They came to our party.

[9] Remember that expectations of higher effort mean, *ceteris paribus*, lower expectations of relevance. The level of effort expected in order to achieve further effects may rise in the course of processing so as to make the initially expected level of relevance – or any level of relevance worth the effort – appear unachievable. At this point, a sound relevance-driven cognitive system cuts its losses and turns its attention to some other information available.

By asking a question Peter has shown what would be relevant to him, namely knowing the answer to his question. Mary's reply does not directly answer him. However, since Mary's utterance carries a presumption of relevance, he is encouraged to look for a way in which the information she explicitly gives (that the Smiths had come to their party) might allow him to infer whether or not she wants to go the party at the Smiths. As is often the case with indirect answers to questions, what is needed is an implicit premise from which, together with the explicit content of the utterance, the expected information can be derived as an implicit conclusion. A factor easily evoked in the context is the existence of a social rule of reciprocation. If Mary's utterance succeeds in bringing this social rule to Peter's mind, then he will be entitled to think that she wanted him to remember this rule and see it as relevant in the context. From there, the implicit conclusion that Mary wants to go to the Smiths' party will be easy to infer. Peter might draw many other inferences from Mary's utterance, for instance that she is not too upset by the risk of meeting the Joneses at the Smiths', but since the expected level of relevance has been reached, he has no reason to assume that such further inferences are part of her meaning, or even that they have crossed her mind.

3.2. Expectations of relevance in an experimental setting

It may be asked whether ordinary considerations of relevance play any role in experimental settings such as that of the Selection Task. In such highly artificial circumstances, the information given is not at all relevant to the subjects, apart from the fact that they are willing to perform the task. However, being the willing subject of an experiment is comparable to the ordinary experience of being willing to help an interlocutor by trying to answer her questions. A standard question expresses the interests of the asker, not the askee. This does not mean that a question is irrelevant to the askee: what is relevant to the willing askee is to know where relevance to the asker lies. In order to identify the kind of information that would be relevant to the asker, the askee need not understand through which cognitive effects this information would achieve relevance. If someone stops you in the street and asks you for the time, you just assume that knowing the time would be relevant to the person, though you don't know, and usually don't care, why or how.

The asker need not be entirely explicit in indicating what information would be relevant to her. The phrasing of her questions may, by manipulating effort and suggesting effects, indicate implicitly in which direction relevance is being envisaged, and direct the askee's comprehension and inference in that direction.

In most experiments in the psychology of reasoning, and in particular in Selection Task experiments, the artificiality of the task is so overwhelming as to discourage any but the lowest expectations of relevance, and anything except the most superficial interpretation processes. The comprehension

component of the task seems uniquely governed by considerations of least effort, and no anticipation of effect. In thematic versions, the introductory narrative does often raise expectations of effect. A standard device used in order to enhance and direct the involvement os subjects in the task is to instruct them to adopt the point of view of a character mentioned in the introduction. We have used also, for the same general purpose, another device: the rule is introduced as a claim actually uttered by a character in the narrative. This conversational setting renders the comprehension task more realistic for the subjects, and easier for us to analyze. By itself, of course, this device is not capable of eliciting correct performance, as shown by the controls where it was also used. Nor is this device necessary, as shown by past literature.

3.3. How relevance guides selection

We start from the following five assumptions. In general:

(i) Subjects understand the task as one of selecting potentially relevant evidence for evaluating the truth of the rule (or conformity to the rule, in the deontic versions).
(ii) Subjects envisage evaluating the rule in the only possible way, that is, indirectly, through its observationally testable consequences.
(iii) Inferring some the consequences of any statement is done spontaneously, as part of the process of comprehension, in order to arrive at a relevant-as-expected interpretation.
(iv) Subjects trust their intuitions, that is, the output of their spontaneous inferential abilities; without any further examination, they take the directly testable consequences that they have inferred in comprehending the rule to be the consequences through which the rule should be tested.
(v) Subjects select the cards the observation of which may directly test these spontaneously derived consequences.

3.4. What inferences may be drawn from the rule?

Some of the inferences that can be drawn from general conditional statements are similar to those that can be drawn from particular conditional statements, while others are specific to general conditional statements. Both general and particular conditional statements may provide the major premise of a valid conditional syllogism of either the *modus ponens* or the *modus tollens* type. From general conditional statements, moreover, affirmative or negative existential conjunctive statements may also be inferred. These latter inferences have been ignored in the Selection Task literature. They may, nevertheless, contribute to the relevance of a general conditional statement. Moreover, affirmative existential conjunctive state-

ments are directly verifiable, and negative existential conjunctive statements are directly falsifiable. The derivation of these existential statements from the conditional rule may play, therefore, a crucial role in subjects' performance on the Selection Task.

More specifically, here is the possibility we want to explore. When subjects derive an existential conjunctive implication from the Selection Task rule, they interpret their task as one of identifying the evidence capable of directly testing this implication (and thereby indirectly testing the rule). However, the evidence displayed is incomplete: only one feature is visible on each card, while the other feature is hidden. Observing the presence of a given conjunction of features requires, then, examining the hidden part of the appropriate cards. The appropriate cards are, trivially, those which display either one of the two features mentioned in the existential conjunctive implication. These, then, are the cards that subjects select.

Looking at the conditional rule of the Selection Task in this light, we distinguish three interesting cases.

Case (a): As is established beyond controversy, conditional statements make *modus ponens* inferences particularly available – and much more available than *modus tollens* inferences (see Evans et al., 1993, chap. 2, and references therein). In the case of a particular conditional statement of the form [P→Q], a minor premise of form [P] must be provided to allow the inference of [Q]. In the case of general conditional statement of form [$\forall x(Px \rightarrow Qx)$], a minor premise of form [Pa] must be provided to allow the inference of [Qa]. In other terms, general conditional statements may typically achieve relevance by allowing one to infer of any item encountered and observed to have the feature P that it also has the feature Q.

In the Selection Task, this way of achieving relevance will lead to the selection of the P card, about which it yields an observationally testable inference, namely that this card also has the feature Q.

When the rule is interpreted "biconditionally" (i.e., as implying its converse [$\forall x(Qx \rightarrow Px)$]), an interpretation often licensed by the context, testing the rule through its *modus ponens* consequences will lead to the selection of both the P and the Q cards. For lack of space we will not consider such cases of biconditional interpretation here (but see Cara & Broadbent, 1995). We want to argue however, that the selection of the P and Q cards need not always be based on a biconditional interpretation. This is illustrated by many past experiments where the thematic content made a biconditional interpretation implausible and where a significant proportion of P and Q selections was nevertheless recorded (e.g. Manktelow & Evans, 1979; see also Experiment 2 below). We explain these selections in the next paragraph.

Case (b): In formal logic, a universally quantified conditional statement of

the form (19) does not entail the existentially quantified statement of the form (20):

(19) [$\forall x$ ($Px \rightarrow Qx$)] [for any item x, if x has the feature *P*, x has the feature *Q*]
(20) [$\exists x$ (*P*x and *Q*x)] [There are items x that have the feature *P* and the feature *Q*]

In other words, a general conditional statement may be true, even though it does not have any instances (i.e., even though there are no items having the feature *P*).

However, in ordinary linguistic communication, it is very common for general conditional statements to be understood as implying that there are positive instances, that is, that there are *P* cases of which the statement asserts that they all also have the feature *Q*. This can be accounted for either semantically, or pragmatically. A semantic account would consist in the claim that, in ordinary language, "all" implies "some" (Strawson, 1952, chap. 6; for empirical evidence, see Begg & Harris, 1982; Newstead, 1989; Newstead & Griggs, 1973; Politzer, 1990), and that this extends to the implicit universal quantifier of the general conditional. A pragmatic account would consist in the claim that, in most contexts, a general conditional statement is irrelevant unless it has instances. When this is so, a general conditional statement is presumed to have instances. In other words, given the presumption of relevance, most statements of form (19) pragmatically imply a proposition of form (20).

In a standard Selection Task, in particular, subjects may be told that there are letters on one side of the cards and numbers on the other side. This raises – but does not answer – the question: what combinations of letters and numbers may be found? Being told, say, that if there is an A on one side of a card, then there is a 7 on the other side, allows subjects to infer, given the presumption of relevance, that there are cards that combine an A and a 7. This may contribute to satisfying, or even satisfy their low expectations of relevance. Generally speaking, subjects may derive from the universally quantified conditional rule the existential assumption that there are cards which have both the *P* and the *Q* features. This will lead to the selection of the two cards that may singly verify or jointly falsify this existential assumption, that is, the *P* card and the *Q* card.

Of course, a bit of meta-inferential thinking would show that if the *P* card fails to verify the existential assumption that there are *P-and-Q* cards, its being verified by the *Q* card would be irrelevant, since the rule would be falsified anyhow, through one of its *modus ponens* consequences. *P* and *Q* cards selections tend to show that subjects don't engage in meta-inferential, critical thinking. If our account is correct, these unsound *P* and *Q* cards selections may nevertheless draw on a sound logico-pragmatic inference,

from a universally quantified conditional to an existentially quantified conjunction.

Case (c): A universally quantified conditional statement of the form (21) entails a negative existentially quantified statement of the form (22),[10] or, equivalently, it contradicts a positive existentially quantified statement of the form (23):

(21) [$\forall$x (Px$\rightarrow$ Qx)] [for any item x, if x has the feature P, x has the feature Q]
(22) [not-($\exists$x (Px and *not-Q*x))] [There are no items x that have the feature P and not the feature Q] or, equivalently, [There are no items x that have the feature P and the feature *not-Q*]
(23) [$\exists$x (Px and *not-Q*x)] [There are items x that have the feature P and not the feature Q] or, equivalently, [There are items x that have the feature P and the feature *not-Q*]

Though entailing (22) and contradicting (23) are logically equivalent, they are not computationally or representationally identical. Deducing a negative existential statement such as (22) from a general conditional statement such as (21) involves the ability to derive implications and that of representing explicit negations of complex propositions. Treating the relationship between (21) and a positive existential statement such as (23) as one of contradictions involves the ability to modify the course of an inferential process when it encounters a contradiction, and that of representing the implicit negation contained in the very idea of contradiction.

Similarly, though [not having the feature Q] and [having the feature *not-Q*] may be logically equivalent predicates, they are not computationally or representationally identical. The former involves representing by means of an explicit negation the absence of a feature, while the latter may involve representing a feature that merely implies the negation of another one.

All these representational and computational abilities are clearly present in humans, but they may differ in cost. Judging false a positive statement seems to be easier than judging true an explicitly negative one, and, generally, implicit negations seem to be psychologically easier to represent than explicit negations (see Braine, Reiser, & Rumain 1984, pp. 343–345; and Horn, 1989, chap. 3 for reviews of the vast relevant literature).

In relatively rare, but not exceptional circumstances, general conditional statements achieve relevance by contradicting a positive existential statement or by implying a negative existential statement. Here is a double conversational example.

[10] Of course, (25) not only entails (26), but it is entailed by it. The two forms are logically equivalent, and this is why testing the rule through this directly testable entailment is a logically correct method. There is no reason to assume, however, that subjects who spontaneously infer (26) from (25) in some contexts are aware of this equivalence.

Consider a situation where an existential assumption of form (23), that there are items combining the feature *P* and the feature *not-Q*, is being either believed or wondered about. For instance, Peter might believe that there is cheap champagne, or he might wonder whether there is cheap champagne. Should he express or imply this belief as in (24), or should he express this interrogation as in (25), Mary might answer by means of the general conditional statement of (26):

(24) *Peter:* Let's buy some cheap champagne!
(25) *Peter:* Is there cheap champagne?
(26) *Mary:* If it is real champagne, then it is expensive

When uttered in reply to (24) Mary's answer (26) achieves relevance through contradicting Peter's positive existential presupposition that there is cheap champagne. When uttered in reply to the question (25), Mary's answer achieves relevance by implying the negative existential proposition that there is no cheap champagne. Note that Mary does not use any explicit negative item. She takes advantage of the existence of a predicate "expensive", which is the contrary of "cheap". This absence of negative items, in spite of Mary's negative intent, makes, we suggest, her utterance marginally easier to process, and, to that extent, more relevant.

Mary's conditional utterance does not achieve relevance in the more common manners described in case **(a)** and case **(b)** above: it neither suggests drawing *modus ponens* inferences about items satisfying the antecedent ("This is a bottle of champagne, therefore it is expensive"), nor deriving the positive existential conclusion that there are items satisfying both the antecedent and the consequent ("There are expensive champagnes"). Mary's utterance achieves relevance by contradicting Peter's presupposition or by answering his existential question in the negative. Mary could have conveyed the same information in other ways, for instance by means of a negative existential statement: "There is no cheap champagne", or by means of a positive universal statement: "Champagne is always expensive." However, the conditional form Mary uses suggests something more: that the fact that champagne is expensive is somehow inferable from the fact that it is champagne. This added cognitive effect justifies the extra effort imposed by the conditional form, and makes her utterance pragmatically felicitous.

Imagine now a Selection Task where the search for relevance will cause subjects to see the rule as a denial of there being *P-and-(not-Q)* cases, or as an assertion that there are no *P-and-(not-Q)* cases. If subjects imagine testing the rule through either of these particular consequences, they will select the *P* and the *not-Q* cards, the observation of which may directly test the truth or falsity of the proposition denied. This is, of course, the correct selection. Notice that it is arrived at not by applying a *modus tollens* schema to the *not-Q* card, but by exploiting in a non-reflective, spontaneous manner

a logical equivalence between quantified formulas. Notice also that subjects who follow this path do search for falsificatory evidence, but this reflects not a Popperian disposition to falsify, but the mere fact that the rule itself is interpreted as a denial or as a implicitly negative statement.

For reasons of space, we will focus almost exclusively on cases of denials. What we say can be extended in an obvious way to cases of negative existential implications.

3.5. *Explaining and predicting*

We have described three different ways in which a general conditional statement may achieve relevance. These are not the only ways in which such a conditional statement may be found relevant. For instance, case **(a)** and case **(b)** may combine yielding a *P* and *Q* cards selection with double reasons for the selection of the *P* card. Case **(b)** and case **(c)** may combine, favoring the selection of the *P*, *Q* and *not-Q* cards.

As we already mentioned, considerations of relevance may favor a "biconditional" interpretation of a general conditional statement. If, *mutatis mutandis*, a biconditionally interpreted conditional rule is seen as relevant in a manner parallel to case **(a)** or to case **(b)**, this will favor the selection of the *P* and the *Q* cards, with the same effects as in case **(b)**, though for different reasons. If, *mutatis mutandis*, a biconditionally interpreted conditional rule is seen as relevant in a manner parallel to case **(c)**, this will favor the selection of the four cards.

When discussing recent work on deontic versions of the task, we will consider yet another way in which relevance may be sought, yielding another pattern of selection (that of the *not-P* and of the *Q* cards). Our goal in this paper is not to explore all the possible cases, but to show, more generally, that the ways in which a conditional rule is likely to be found relevant are predictable from the content and the context of the statement, and that this in turn will predict subjects' selections. For this purpose, exploiting the three cases we have discussed should suffice.

The two most common ways of achieving relevance, that is, cases **(a)** and **(b)**, yield the most commonly found pattern of selection, that of the *P* card alone or that of the *P* and *Q* cards. Case **(c)** stands apart because of its relative rarity, and, more importantly, because it is the only case which yields the fully correct selection pattern of the *P* and the *not-Q* cards.

3.6. *A recipe for constructing easy versions of the task*

Researchers have been at pain to produce such correct selection patterns. As we mentioned, they have succeeded in doing so in a reliable and robust manner only in the case of "if P, then not-Q" abstract rules, and in the case of deontic rules (or else in task inspired by the Selection Task proper, but differing from it in crucial respects). We will argue, in the concluding

section, that negative-consequent abstract rules are construed as denials, essentially because of relevance considerations on the effort side. We will argue that deontic versions elicit correct responses because they achieve relevance through implicitly forbidding the occurrence of cases of *P-and-(not-Q)* (forbidding occurrences being here the deontic counterpart of denying occurrences). We will argue that these and all successful versions of the task are instances of case (c). If we are right in our claim, we should be able to do more than reinterpret previous findings. We should be able to spell out a recipe for producing versions of the task that elicit a majority of correct selections, and to demonstrate it. To this aim, we now turn.

The basic idea should now be clear. In order to build an easy Selection Task, one must cause subjects to interpret the rule as a denial of the occurrence of *P-and-(not-Q)* cases.[11] In order to achieve this result, two difficulties must be overcome: one on the effort side, and the other on the effect side.

On the effort side, interpreting the rule as a denying [there are *P-and (not-Q)* cases] is more effort-demanding than representing it as implying [there are *P-and-Q* cases]. The former involves two negations, an implicit one in the propositional attitude of denial, and an explicit or implicit one in the proposition denied, while the latter involves no negations at all. The relative difficulty of representing and processing negative information is well established (Horn, 1989; Wason, 1959). The greater effort that is required to arrive at an interpretation of the rule as a denial makes this interpretation less likely than the standard interpretation for two reasons: for the trivial psychological reason that whatever requires more effort is less likely to be arrived at, and also, if Relevance Theory is right, because relative effort is used as evidence of intended interpretation, favoring less effort-demanding alternatives.

On the effect side, general conditional statements create expectations of cognitive effects achieved by deriving conclusions of the form [*Q*a] from the rule together with premises of the form [*P*a], or achieved through deriving a conclusion of the form [there are *P-and-Q* cases] from the rule together with the presumption of relevance. These are the standard ways in which general conditional statements achieve relevance, and in the absence of evidence pointing in another direction, these are the expectations of relevance to which they give rise. The first of these expectations (that of *modus ponens* effects) is quite compatible with a correct selection. The second of these expectations (that of conjunctive existential effects), however, is not compatible with a correct selection, and it must somehow be avoided.

In order to prevent subjects from interpreting the conditional rule in the most common way, it is necessary to act both on the effort and on the effect

[11] Or, as an implicit statement of the absence of *P-and-(not-Q)* cases, or, in the deontic versions, as a rule forbidding of the occurrence of such cases.

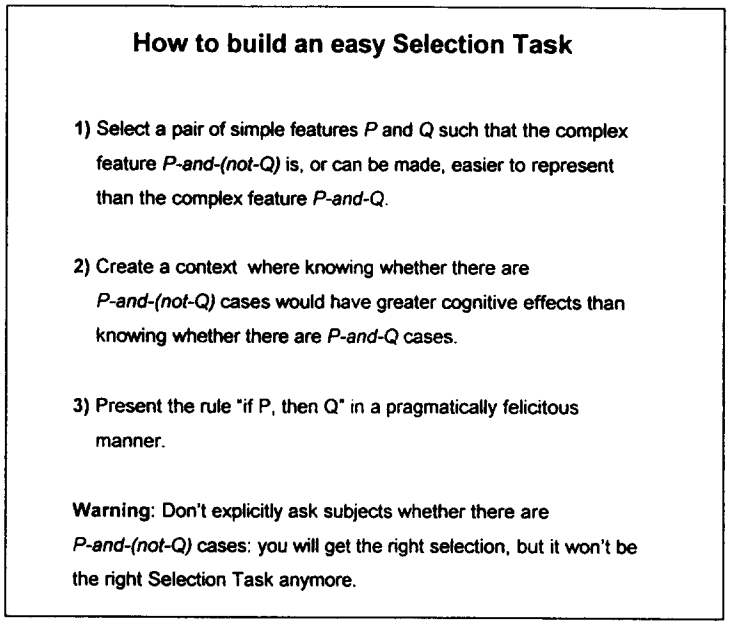

Fig. 4. The recipe.

side. The basic recipe is presented in Fig. 4. Here we comment on it in detail.

On the effort side, one must select a pair of features P and Q such that the complex feature *P-and-(not-Q)* is as easy, or preferably easier to represent than the complex feature *P-and-Q*. For this, one may, for instance:

- Introduce and render salient a *P-and-(not-Q)* category in the text of the problem, for example, the category of "virgin-mothers" (where P = with children, and Q = having had sex) that we used in Experiment 1.
- Take advantage of a lexicalized concept the meaning of which combines the presence of a feature P and the absence of another feature Q, for example, "bachelor" (where P = male, and Q = married) that we used in Experiment 2.
- Give subjects an intermediate task where they have to single out the *P-and-(not-Q)* cases, as we did in several pilot experiments.
- Define a structurally simple universe where the complementary of P and that of Q are two explicitly introduced positive features R *(= not-P)* and S *(= not-Q)*, so that there are four possible combinations – *P-and-Q*, *P-and-S*, *R-and-Q*, and *R-and-S* – all equally easy to represent, as we did in Experiment 4.

On the effect side, one must invent a context where knowing whether there are *P-and-(not-Q)* cases would have equal or preferably greater cognitive effects than knowing whether there are *P-and-Q* cases. For this, one may, for instance:

- Make the existence of *P-and-Q* cases relatively trivial, and the possible existence of *P-and-(not-Q)* cases contentious, as we did in Experiment 1 and 2.
- Make the existence of *P-and-(not-Q)* cases diagnostic of something that matters in the context, as we did in Experiments 3 and 4.
- Make the existence of *P-and-(not-Q)* cases undesirable from a point of view that the subjects are instructed to adopt, as is done in most deontic versions.

For these pragmatic manipulation to be fully effective, the rule "if P, then Q" should be introduced in a pragmatically felicitous manner. Otherwise, the blatant artificiality of the task is likely to refocus subjects' attention on the question: What does experimenter expect from us? In other terms, the rule should carry cognitive effects, and do so without reminding subjects of the availability of some alternative linguistic form that would have achieved the same effects with much less effort. A modicum of pragmatic felicity can be achieved by having, for instance, the rule introduced as a statement made at an appropriate juncture by a character as we did in all our experiments.

The conditional form "if P, then Q" is more likely to be a felicitous way of conveying that there are no *P-and-(not-Q)* cases, when the fact that an item has the feature *Q* is inferable from the fact that it has the feature *P* (otherwise, why not just say "There are no *P-and-(not-Q)*s"?). For instance, the fact that a woman has had sex is inferable from the fact that she has children, and this makes the form "if a woman has children, then she has had sex" a felicitous way of denying the existence of virgin-mothers (Experiment 1). This particular feature cannot be exploited in abstract version, where the relationship between what is on one side of the card, and what is on the other, is intrinsically arbitrary. There a certain degree of manifest artificiality seems unavoidable, but not radically detrimental.

Generally, if one follows the recipe strictly and carefully one can produce easy versions of the Selection Task in any cognitive domain. Following the recipe carefully does require a bit of imagination and tact. There is a risk, in particular, of introducing unintended pragmatic twists which cause subjects to seek relevance in a unforeseen direction and to select cards accordingly. If such a pragmatic twist is indeed what causes such unforeseen results, then an independently justifiable and testable pragmatic explanation should be able to account retroactively for them. Otherwise, we found that even clumsy following of the recipe will produce at least a clear tendency in the expected direction, a tendency which can be turned into a significant result by improving the design in stricter accordance with the recipe.

EXPERIMENTS

In the four experiments we present in this section, we compare two conditions (plus two intermediate ones in Experiment 4). In one type of condition, which we will call the "relevance condition," we expect subjects to understand the problem in a way such that *P-and-(not-Q)* cases will appear relevant to them. In another type of condition, which we will call the "irrelevance condition", we expect that subjects' understanding of the problem will not make *P-and-(not-Q)* cases appear relevant. (We do *not* mean that, in the irrelevance condition, the problem itself will appear irrelevant to subjects.)

Experiments 1 and 3 were conducted in France and in French. Experiments 2 and 4 were conducted in Italy and in Italian.

EXPERIMENT 1: THE VIRGIN-MOTHERS PROBLEM

The main aim of this first experiment was illustrative. We wanted to show concretely how to build a "relevance" version of the task by applying the recipe described in the previous section, and to contrast performance on such a version with performance on a standard abstract version of the task, in a within-subjects design.

The scenario we created involved an imaginary Californian sect, the Haré Mantra (see Fig. 5). The cards represented women members of the sect. The two features *P* and *Q* we selected, in accordance with the first step of the recipe, were *has a child (P)* and *has had sex (Q)*. The rule is therefore, "If a woman has a child, then she has had sex." The counter-example to that rule would be provided by the case of a woman who has a child but has not had sex.

The feature *P* (*has a child*) is lexicalized in English as "mother" (for a woman), while its contrary is not lexicalized. The feature *Q* (*has had sex*) is not lexicalized in English, while its contrary, *not-Q*, is lexicalized as a positive feature, "virgin". (The same holds in French, the language in which we conducted this experiment.) Of course, it is generally taken for granted, for empirical rather than definitional reasons, that "mother" refers to a woman who, having a child, has had sex, while "virgin" refers to a person who, not having had sex, does not have a child. However, we reasoned that, if we created a context where the possibility of virgin-motherhood was highlighted, the automatic inference from mother to non-virgin, and from virgin to non-mother would be suspended. In such a context, creating the lexical compound "virgin-mother" would make the *P-and-(not-Q)* combination (*has a child and has not had sex*), that is, the counter-example to the rule, as easy as, or easier to represent than the three other combinations, *P-and-Q* (*has a child and has had sex*), *(not-P)-and-Q* (*does not have a child*

Until recently, it was obvious that a woman who has children has had sex. With artificial insemination, it is now possible for a virgin to have children. The leader of the Haré Mantra (a very secret religious, Californian sect) has been accused of having had some of his sect's virgin girls artificially inseminated.

His goal, it is claimed, is to create an elite of "Virgin-Mothers" alleged to play a particular role in his religion. The head of the Haré Mantra makes a joke out of these allegations. He claims that the women of his sect are, without exception, like any other women : **if a woman has a child, she has had sex**.

Imagine that you are a journalist and that you are preparing an article on the Haré Mantra. You learn that a gynecological survey has been carried out among the Haré Mantra women. Some of them might be "Virgin Mothers". You go and visit the doctor who carried out the gynecological survey. Unfortunately, the doctor pleads professional secrecy and refuses to tell you what he discovered.

You realise that, before you on the doctor's desk, there are four individual information cards about Haré Mantra women examined in the gynecological survey. However, these four cards are partially concealed by other papers (as shown below). Of two cards, one can only see the top where you can read whether the woman has children or not. Of the other two cards, you can only see the bottom where you see whether the woman has had sex or not. You are determined to take advantage of a moment in which the doctor turns his back to uncover the papers and to learn more .

Indicate (by circling them) those cards that you should uncover in order to find out whether what the leader of the Haré Mantra says (**"if a woman has a child, she has had sex"**) is true, as far as these four women are concerned. Indicate only those cards that it would be absolutely necessary to uncover.

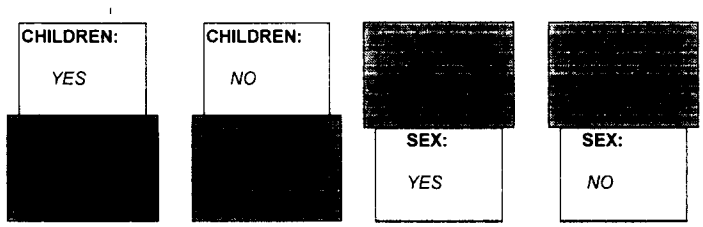

Fig. 5. The virgin-mothers problem.

and has had sex), *(not-P)-and-(not-Q)* (*does not have a child and has not had sex*). Thus we hoped to fulfill the first step of our recipe.

On the effect side, the presence of women who, having a child, have had sex is generally too trivial to be of relevance, while the presence of virgin-mothers, if it were plausible, would be highly relevant (i.e., rich in implications). All it takes to exploit such contrasted expectations of relevance based on general knowledge, then, is to create a context where virgin-motherhood is indeed plausible but not certain. For this, we had the leader of the sect suspected of creating an élite of virgin-mothers by means of artificial insemination. Moreover, we instructed subjects to adopt the point of view of a journalist investigating these allegations, and for whom, therefore, the cognitive effects of finding out that there were virgin-mothers would be much greater than finding out that there were ordinary mothers

among the women of the sect. Thus we were following the second step of the recipe.

The third step of the recipe is: present the rule "if P, then Q" in a pragmatically felicitous manner. That is, the information provided by the rule must be relevant at the juncture at which it occurs, and its being presented in a conditional form must contribute to, rather than deter from, its relevance. For this, we had the rule introduced as a statement of the leader in response to the allegation that there are virgin-mothers in this sect. In such a context, we reasoned, the statement "If a woman has a child, then she has had sex" achieves relevance as a denial of the allegation. Moreover the conditional form is appropriate because, generally, the fact that a woman has had sex is indeed inferable from the knowledge that she has a child. By using the conditional form rather than resorting to a flat denial, the sect leader is effectively appealing to his audience's commonsense knowledge.

Having thus followed the recipe, we predicted that, in the virgin-mothers task, selection of the *p* and *not-Q* cards would be the dominant pattern, in contrast to a standard task administered to the same subjects.

4. Method

4.1. Subjects

Twenty-seven computer science undergraduate students of the University of Paris VI participated as subjects. They were randomly assigned to one of two groups, depending on the order in which they received the two tasks: standard task–virgin-mothers task ($n = 13$); virgin-mothers task–standard task ($n = 14$).

4.2. Procedure and materials

We ran the experiment in a lecture hall. Each subject received a three-page booklet. The first page explained that the task was about reasoning and gave general task instructions. The second and third pages presented the two Selection Tasks in the standard and in the new, virgin-mothers versions. Subjects were told to read the problem carefully and were informed that they did not have any time limit. The virgin-mothers version is presented in Fig. 7.

The standard task was as follows: "Imagine a set of cards each with a letter on one side and a number on the other side. Four of these cards are represented below: two on the letter side and the other two on the number side." Pictures of the four cards followed, showing respectively the letters "E" and "M", and the numbers "7" and "4".

Subjects were asked to indicate which cards they would need to turn over

in order to find out whether the following statement is true or false: "If there is a E on one side of the card, then there is a 7 on the other side of the card."[12]

5. Results

Table 1 presents the main results. The *P* and *not-Q* cards combination were selected by 78% of the subjects in the relevance condition against 26% in the irrelevance condition. We computed a McNemar test to evaluate the significance of the change of each subject's response to the two versions of the Selection Task. The test showed that significantly more subjects selected the *P* and *not-Q* cards in the virgin-mothers task than in the standard task ($\chi^2(1) = 12.07$, $p < .001$). The figure of 26% for the standard task is, nevertheless, rather high, compared to the usual 10% or so found on this task. This suggests a transfer effect. The tests indicated that a positive transfer did not occur between tasks. When the virgin-mothers task was presented first, more subjects selected the P and the not-Q cards in the standard task ($z = 2.04$, $p < .01$; rank-sum analysis for frequency tables, specific tests; this analysis was chosen over more traditional or non-directional ones because it makes it possible to separate specific and non-specific hypothesis testing, cf. Meddis, 1984).

These results clearly confirmed our prediction. Faced with both a relevance and an irrelevance version of the Selection Task, the same subjects produced radically different responses: correct ones in the first case, and incorrect ones in the second. Our explanation of this difference is this. In the standard task, subjects find little relevance in the rule, and the little relevance they find has to do with the fact that the rule allows them to infer that the card that has the feature *P* also has the feature *Q*, and that there

Table 1
Frequencies of selection patterns in the two conditions (Experiment 1, virgin-mothers)

	Condition	
Pattern	Relevance	Irrelevance
P, not-Q	21	7
P, Q	1	7
P	3	9
Q	0	2
P, Q, not-Q	1	1
P, not-P, Q	0	1
not-P, not-Q	1	0

$n = 27$ in both conditions.

[12] In this, as well as all the conditions of our four experiments, for the sake of parallelism, we presented the cards in the same formal order: *P, not-P, Q, not-Q*.

are *P-and-Q* cards. In the virgin-mothers task the rule is given a relevant interpretation as a denial of the existence of *P-and-(not-Q)* cases. In both tasks, subjects choose the cards that provide a direct test of what they see the rule as implying or as denying.

Nevertheless, there are many differences between the two conditions tested: one is "abstract", the other is concrete, one is much longer than the other, only one gives a rationale for selecting only the necessary evidence, only one requires the subjects to adopt a specific perspective, only one presents the rule as an actual utterance, etc. In order to provide much stronger corroboration of the relevance-theoretic explanation of the Selection Task, we want a much better control of the factors capable of affecting performance. More specifically we want the relevance and the irrelevance conditions to differ as little as possible so as to help us single out the factors responsible for differences in performance. This can be achieved, in particular, by keeping the context as similar as possible across conditions while changing the rule, as we have done in Experiments 2 and 3, or, on the contrary, by keeping the rule identical across conditions while changing the context, as we have done in Experiment 4.

EXPERIMENT 2: THE BACHELORS PROBLEM

The main aim of Experiment 2 is to show that two similar (and equally thematic) context-rule combinations produce very different performances as a function of the way in which the rule achieves relevance in the context.

The scenario used in both the relevance and the irrelevance condition of this experiment concerned the recruitment of volunteers to take care of children in an exchange program (see Fig. 6). The cards represented individual volunteers. The feature *P* was that of being male, in both conditions. The feature *Q* was that of being married in the relevance condition, and that of being dark-haired in the irrelevance condition. In the case of the relevance condition, the *P-and-(not-Q)* counter-example to the rule is lexicalized as "bachelor" in English, and as "scapolo" in Italian (the language in which the experiment was conducted). The existence of a lexicalization, particularly when the word is a common one, guarantees, we assume, that the counter-example to the rule is relatively easy to represent (though we don't mean to imply that lexicalization of the counter-example is necessary – see Experiments 4 and 5). In the case of the irrelevance condition, the *P-and-(not-Q)* counter-example, that is, that of a man who does not have dark hair, is not lexicalized in Italian, nor is it particularly salient either in the culture, or in the scenario. We assumed therefore that it would be harder to represent the counter-example in the irrelevance condition than in the relevance condition.

The rule is introduced by a character in a discussion. In the relevance condition, she expresses doubts about the willingness of men, and in

As a part of a program of cultural exchange with the English town of Milton Keynes, the town of Padua is expecting the visit of a large group of English school children. The City Council of Padua has asked for volunteers to help take care of these children and make sure that they have a good time. The volunteers have to fill very detailed cards.
Mr Rossi and Mrs Bianchi, two clerks of the City Council, are about to sort the double-sided cards which have been filled by the volunteers.

[Relevance condition]

Mrs Bianchi, who has strong views on many things, says: "I'm sure there will be only women volunteers. Men don't want to take care of children!". However, Mr Rossi assures her: "You are wrong, there are also male volunteers!". "Really? That is remarkable, says Mrs Bianchi. Still, bachelors, at least, won't be at all interested in children! I bet you, if a volunteer is male, then he is married".

[Irrelevance condition]

Mrs Bianchi, who has strong views on many things, says: "Men with dark hair love children! I bet you, if a volunteer is male, then he is dark-haired".

Mr Rossi replies: "I cannot tell for sure, but I am convinced you are wrong and I accept your bet!".
The cards filled by the volunteers indicate their sex on the front side and their marital status on the back side. Presently, on the table of the two clerks there are four cards. Two of them are with the front side up and the two others with the back side up. It is therefore impossible to find out whether Mrs Bianchi is right without turning over one or more of the cards.
Your task is to decide which cards it is absolutely necessary for Mr Rossi to turn over in order to make sure whether, as far as these four volunteers are concerned, Mrs Bianchi is right or wrong in betting that

[Relevance condition]

If a volunteer is male, then he is married

[Irrelevance condition]

If a volunteer is male, then he is dark-haired

Please circle only those cards that you think must be turned over and justify your choice below.

Fig. 6. The bachelors problem.

particular unmarried men, to take care of children, and then bets that, "If a volunteer is male, then he is married." In this context, the most salient cognitive effect of the conditional statement is to cast doubt on the presence of bachelors among the volunteers. In the irrelevance condition, the rule is introduced by a character who expresses the view that dark-haired men love children and then bets: "If a volunteer is male, then he is dark-haired." The most salient cognitive effect of the conditional statement, in this case, is to suggest that one should definitely expect to find dark-haired men among the volunteers.

For reasons having to do both with effort and with effect, then, we predicted that the rule wold be interpreted as a denial in the relevance condition and not in the irrelevance condition, and that as a result, significantly more subjects would make the correct selection in the relevance condition than in the irrelevance condition.

6. Method

6.1. Subjects

Thirty-six undergraduate students studying psychology at the University of Trieste participated as subjects. They were randomly assigned to one of two conditions (relevance condition, $n = 17$; and irrelevance condition, $n = 19$).

6.2. Procedure and materials

Subjects were run in two large groups for each condition. A two-page booklet was used. General task instructions were given on the first page. Subjects were told to read the question (concerning a reasoning problem) carefully and to take the time they required. For the text of the two versions, see Fig. 6. In the relevance conditions the cards pictured showed: "Gender: Male", "Gender: Female", "Marital status: Married", and "Marital status: Unmarried". In the irrelevance condition, they showed: "Gender: Male", "Gender: Female", "Hair color: Black", and "Hair color: Fair".

7. Results

The results are summarized in Table 2. They clearly support our predictions. While only 16% of the subjects correctly solved the selection problem in the irrelevance condition, a significantly higher percentage of subjects (65%) produced the correct solution in the relevance condition ($z = 2.96$, $p < .01$; rank-sum analysis for frequency tables, specific tests; cf. Meddis, 1984).

As predicted, two versions of the same general scenario produce a

Table 2
Frequencies of selection patterns in the two conditions (Experiment 2, bachelors)

	Condition	
Pattern	Relevance ($n = 17$)	Irrelevance ($n = 19$)
P, not-Q	11	3
P, Q	2	10
not-P	1	2
P, Q, not-Q	2	0
All	0	2
Q	1	0
not-Q	0	1
P, not-P, not-Q	0	1

different selection performance as a function of the pragmatic appropriateness of interpreting the conditional rule as a denial of the occurrence of *P-and-(not-Q)* cases. Note that the relevance-theoretic account predicts not only the good performance in the relevance condition, but also the specific performance elicited in the irrelevance condition. In the irrelevance condition, over 50% of the subjects selected the *P* and the *Q* cards. The usual explanation of *P* and *Q* cards selection as resulting from a biconditional interpretation is not plausible here: nothing whatsoever in the narrative or in the background knowledge of subjects would invite them to assume that, if a volunteer is dark-haired, then he is male, rather than female. On the other hand, the assertion, "If a volunteer is male, then he is dark-haired" is presented by Mrs. Bianchi as an illustration of her claim that "men with dark hair love children!" To achieve relevance in this context, the conditional assertion must be interpreted as implying that there are (significantly many) dark-haired men (i.e., *P-and-Q*s) among the volunteers.

These results are also of interest as counter-evidence to two recent theoretical proposals (Kirby, 1994; Oaksford & Chater, 1994) regarding the Selection Task. According to Oaksford and Chater (1994) "card selection frequencies are a monotonic function of the expected information gain associated with each card". In particular, they calculate that the informativeness of the *not-Q* card increases with the probability of occurrence of *P* cards. Similarly, according to Kirby (1994) the proportion of subjects selecting the *not-Q* card should increase with the ratio of *P* cases to *not-P* cases.

In the relevance condition of Experiment 2, Mrs. Bianchi explains why males are unlikely to volunteer. In the irrelevance condition no comparable information is given to fix the prior odds that a volunteer is male. In the relevance condition, therefore, the probability *p(P)* of a volunteer being male should be seen as, if anything, lower than in the irrelevance condition, and the probability *p(not-P)* of a volunteer bing female should be seen as correspondingly greater. As we just saw, according to Kirby (1994) and to Oaksford and Chater (1994), this should cause relatively more subjects to select the *Q* card in the relevance condition, and the *not-Q* card in the irrelevance condition. However, the opposite pattern of selection was found: the *not-Q* card was selected by 76% of subjects in the relevance condition versus 31% in the irrelevance condition, while the *Q* card was selected by 29% of subjects in the relevance condition versus 58% in the irrelevance condition.

EXPERIMENT 3: THE UNEMPLOYMENT PROBLEM

The disparity between the instructions in the relevance and the irrelevance conditions of Experiment 1 was very much reduced in Experiment 2. Still, even there, not only the rule, but also some of the information given in

the narrative differed in the two conditions. It might be argued that the two conditions made available different background knowledge – related to marriage in one case, to hair color in the other – thus activating different inferential resources. Therefore, the aim of Experiment 3 was to confirm the results obtained in the previous experiment, while closing any remaining information gap between the two conditions.

In both conditions, the scenario (see Fig. 7) contained exactly the same information. In fact, it was word for word identical, except, of course, for the rule itself, and for the words introducing the rule. The rule is introduced with the words: "For instance, the Prince asserts" in the relevance condition, versus "Of course, the Prince adds" in the irrelevance condition. This is a difference not in informational content, but, strictly, in pragmatic indication. In relevance-theoretic terms, phrases like "of course" and "for instance", or verbs like "assert" or "add" suggest the manner in which relevance is to be achieved (Blakemore, 1987, 1992; Blass, 1990; Ifantidou-Trouki, 1993; Moeschler, 1989; Wilson & Sperber, 1993). Thus, if this difference were a significant factor in subjects' performance – which, in this case, we doubt and have not independently tested – it would provide extra evidence for a pragmatic account of the Selection Task such as ours. Our main motivation, however, in adopting different wordings at this juncture, was just to make the introduction of the rule pragmatically felicitous in both cases.

Imagine that you are a journalist and that you are preparing a piece on the small principality of Bagurstan. The reigning Prince studied Economics at Oxford and has imposed a radical form of liberalism upon his country. In Bagurstan, people retire at 65, students get a salary, but they have no social security, no right to work, no unemployment benefits, no minimum salary, no maternity leave. Yet, the Prince claims that in Bagurstan there are no serious economic or social problems. Economic mechanisms alone allow everyone to find a satisfactory solution.

[relevance condition]	*[irrelevance condition]*
For instance, the Prince asserts, in my country, **if a person is of working age, then this person has a job**	Of course, the Prince adds, in my country, **if a person is older than 65, then this person is without a job**

Before you, on the desk of the Bagurstani civil servant that you are interviewing, are four information cards corresponding to four Bagurstani subjects. The top part of each card indicates the age. The bottom part indicates whether the person has a job. Unfortunately, some papers conceal the bottom of two of these cards, as well the top of two other cards.

Which cards should be uncovered to determine whether what the Prince says

[relevance condition]	*[irrelevance condition]*
(if a person is of working age, then this person has a job)	**(if a person is older than 65, then this person is without a job)**

is true, at least as far as these four Bagurstani are concerned.

Fig. 7. The unemployment problem.

In order to minimize the difference between the two rules, and to avoid their evoking different world knowledge, we used complementary features in the relevance and in the irrelevance conditions. The cards represented adult citizens of an imaginary country, the Principality of Bagurstan. The feature *P* was that of being of working age in the relevance condition, and that of being past working age (above 65, as explained in the text) in the irrelevance condition. The feature *Q* was that of having a job in the relevance condition, and that of being without a job in the irrelevance condition. The rule was therefore, in the relevance condition, "If a person is of working age, then this person has a job", and, in the irrelevance condition "If a person is older than 65, then this person is without a job". Given the complementarity of the rules in the two conditions, exactly the same cards could be used in both.

The *P-and-(not-Q)* counter-example in the relevance condition is that of a person of working age and without a job. This is lexicalized as "unemployed", or "chômeur" in French, the language in which the experiment was conducted, a word of high social salience at the time of the experiment. In the irrelevance condition, the *P-and-(not-Q)* counter-example is that of a person older than 65 and with a job. This is a non-lexicalized category of no particular salience. Thus on the effort side, the recipe was properly followed.

On the effect side, the text which introduces the Prince's statement of the rule describes the country in a way that may suggest that it suffers from many social and economical ills. In particular, the indication that "there is no right to work" may cause subjects to wonder about the level of unemployment. However, this line of inference is explicitly blocked by the Prince, who claims that there are no serious economic or social problems in his country. In this context, we assumed that the Prince's utterance "If a person is of working age, then this person has a job" would achieve relevance and be understood as a denial of the existence of unemployment in Bagurstan. Describing the economy of Bagurstan as a strictly free-market economy, with no social rights whatsoever, and in particular no right to work, served also a second purpose. We wanted to make as certain as possible that subjects would not interpret the rule deontically, and not see the task as one of looking for legal or moral violations (as opposed to factual counter-examples). In the context, the Prince's intention in stating "If a person is of working age, then this person has a job" is clearly to make a statement of fact, and cannot sensibly be construed as a statement of rights.

In the irrelevance condition, the statement: "If a person is older than 65, this person is without a job" could not be interpreted as a denial of some previous inference or suspicion. The relevance of this statement in this context is that of an admission (as suggested by the use of "Of course . . ."): old people are without jobs, but since this was already mentioned, the relevance of such a statement has little to do with its content, but is rather meta-communicative. By this admission, the Prince can be understood as

intending to show his intellectual honesty. In such conditions, the explicit propositional content itself is of limited intrinsic relevance, nor does it have any single clear implication determining relevance, and through which the statement should be tested. We predicted therefore a relative dispersal of responses.

8. Method

8.1. Subjects and procedure

Forty undergraduates studying philosophy and humanities at the University of Paris X (Nanterre) served as subjects. They were randomly assigned to one of two equal-sized conditions (*relevance* and *irrelevance*; $n = 20$) and tested with the same procedure as in previous experiments.

8.2. Material

The same problem was used for both conditions. The only difference concerned the rule and its immediate introduction (see Fig. 7). Pictures of the cards were presented: two with the top part visible and "AGE: 32" on one, and "AGE: 79" on the other, and two cards with the bottom part visible and "JOB: YES" on one, "JOB: NO" on the other.

9. Results

The frequency of selection patterns is reported in Table 3. As can be seen, the relevance condition yields significantly more correct selections of the *P* and *not-Q* cards than the irrelevance condition (70% vs. 25%,

Table 3
Frequencies of selection patterns in the two conditions (Experiment 3, unemployment)

	Condition	
Pattern	Relevance	Irrelevance
P, not-Q	14	5
P, Q, not-Q	3	1
P, Q	1	5
P, not-P, Q	1	3
not-Q	0	2
not-P, Q	1	1
P	0	1
not-P, not-Q	0	1
p, not-p, Q, not-Q	0	1

$n = 20$ in both conditions.

respectively; $z = 2.8$, $p < .01$). This result corroborates our hypothesis: the same scenario elicits a high rate of correct performance in a version in which it achieves relevance through implicitly denying the occurrence of *P-and-(not-Q)* cases. In the irrelevance condition, on the contrary, the responses are scattered among several patterns, as we predicted.

In Experiment 3, the difference between the two conditions is minimal. No information is available in one version and not in the other; both rules express a contingency between the same two features, age and employment. The range of factors that might in principle explain the difference in performance between the two conditions is therefore very narrow. Any explanation of the results would have to draw from this range.

A possible explanatory claim – that would not contradict our explanation, but would limit its import – might draw on the fact that, in the three experiments reported so far, the rules expressed plausible realistic relationships, and the cards represented states of affairs in the world. While such thematic rules and cards are known to be insufficient for eliciting correct solutions (see Manktelow & Evans, 1979), a point confirmed by the irrelevance condition in Experiment 2 and 3, one might argue that they are necessary in order to build an easy Selection Task (leaving aside, of course, the case of negative-consequent rules). Experiment 4, where we used exactly the same abstract rule in all conditions, and varied only the context, will show that this is not the case.

Another possible explanatory claim that might come to mind at this stage is this: in the relevance, but not in the control or irrelevance conditions of Experiments 1–3, the counter-example was lexicalized. It might be contemplated that the lexicalization of the counter-example – which we claim is a non-necessary factor on the effort side of relevance – is necessary, or even necessary and sufficient, for the good performance we have found in these thematic descriptive versions of the Selection Task. However, in Experiment 4, the counter-example to the rules tested was not lexicalized.

EXPERIMENT 4: THE MACHINE PROBLEM

In the relevance conditions of Experiments 1–3, we manipulated the two factors of relevance, effect and effort, in order to cause subjects to interpret the rule as a denial of the occurrence of *P-and-(not-Q)* cases. Our doing so was motivated by a relevance-theoretic account of the task, which the results presented so far do corroborate. Our account of relevance is compatible with the idea that manipulating one factor alone, either effect or effort, would already affect expectations of relevance and hence performance on the task. We expect, however, it to be generally necessary to act simultaneously on both the effect and the effort side in order to change the character of the task, from a distinctly hard one to a distinctly easy one. This is because, as we have argued, the standard spontaneous understanding of a

general conditional statement goes strongly against good performance on the Selection Task.

Several experiments in the literature can be interpreted as showing that making *P-and-(not-Q)* easier to represent could increase the rate of correct response (e.g., Hoch & Tschirgi, 1985; Wason & Green, 1984), and this is described as the key factor by Johnson-Laird and Byrne (1991, p. 80). It could be envisaged, then, that our manipulations of the effort factor in all the relevance conditions have been particularly effective, and by themselves account for our results. If this were so, the relevance-theoretic account would merely add irrelevant effect considerations to, and be parasitic on, a more parsimonious effort-based account.

The purpose of this fourth experiment is to take apart the two factors of relevance, effect and effort, to test their respective roles, and to ascertain whether relevance, which combines the effort and the effect factors in a principled manner, is more explanatory than effort (or, for that matter, effect) taken alone. For this, we created not two but four scenarios, varying the effect and the effort factors separately (see Fig. 8). All scenarios involve a machine that manufactures cards with a number on one side and a letter on the other side.

- In the **effect +** / **effort −** condition, that is, the relevance condition, expectations of effect are raised by making conformity with the rule diagnostic of the machine functioning well again after having malfunctioned and having been repaired. Effort is reduced by defining positively the *not-P* and *not-Q* features. More specifically, the introduction defines a universe where the complementary of *P* and of *Q* are two explicitly introduced positive features: *having a 4 on the number side* (=*not-P*) and *having an A on the letter side* (=*not-Q*). The four possible combinations; 6 and E, 6 and A, 4 and E, and 4 and A, are all equally easy to represent.
- In the **effect +** / **effort +** condition, the malfunctioning-repair story is used on the effect side, but on the effort side the *not-P* and *not-Q* features are not positively defined.
- In the **effect −** / **effort −** condition, nothing on the effect side contributes to making *P-and-(not-Q)* cases particularly relevant, but on the effort side the *not-P* and *not-Q* features are positively defined.
- In the **effect −** / **effort +** condition, that is, the irrelevance condition, nothing on the effect side contributes to making *P-and-(not-Q)* cases particularly relevant, and the *not-P* and *not-Q* features are not positively defined.

The two **effect +** conditions on the one hand, and the two **effect −** conditions on the other hand, differ from one another only on the effort side, while the two **effort +** and the two **effort −** conditions differ from one another only on the effect side. Given this, the predictions that follow from the relevance-theoretic account of the task are self-evident: the best

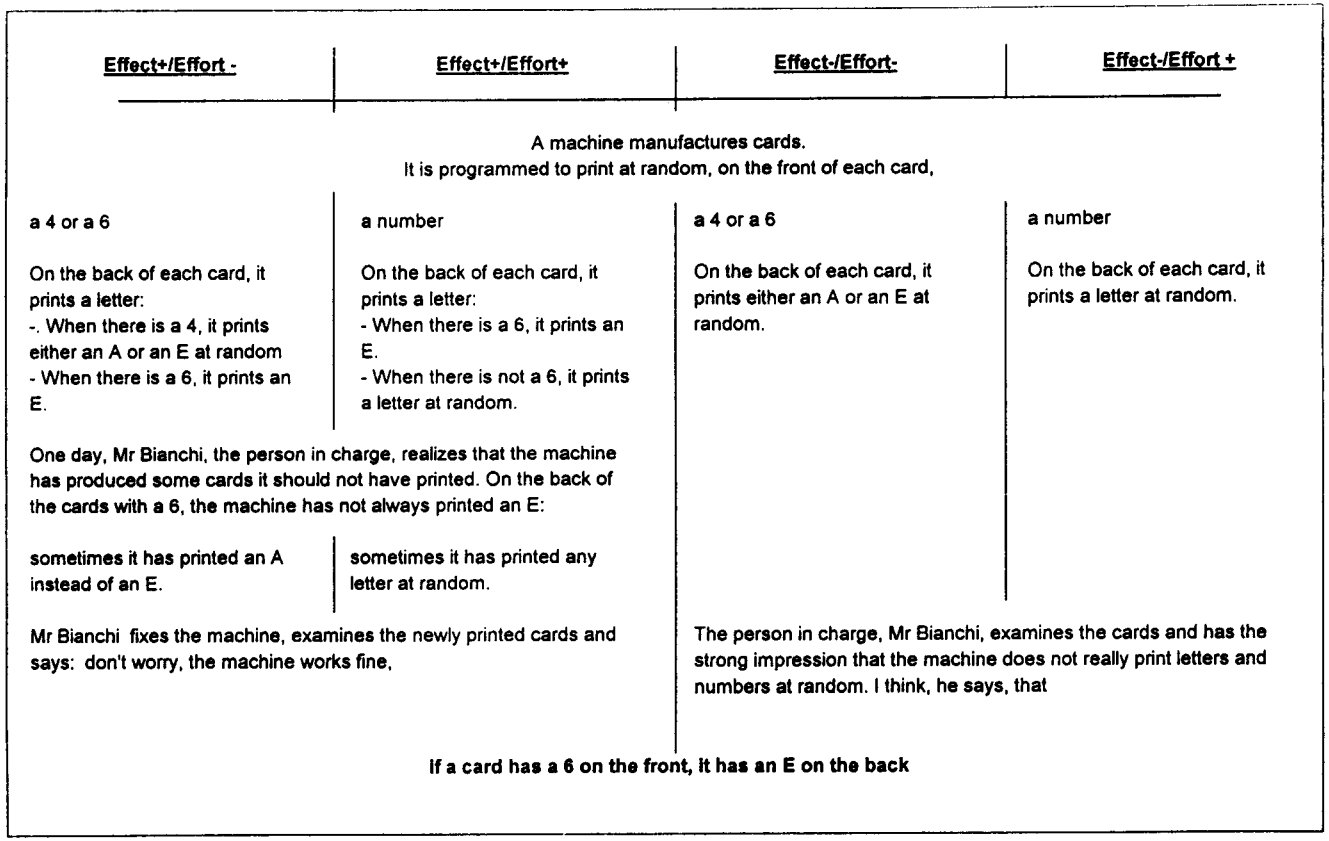

Fig. 8. The four conditions of Experiment 5: the second machine problem.

performance should be with the **effect +** /**effort −** condition, and the worse one with the **effect −** /**effort +** condition. The performance on the **effect +** / **effort +** and on the **effect −** /**effort −** condition should be at an intermediary level between the two other conditions. Moreover, the two factors, effect and effort, should each on its own contribute to good performance.

10. Method

10.1. Subjects

The subjects were eighty-four first-year psychology students at the University of Padua.

10.2. Procedure and materials

The subjects were randomly assigned to one of four conditions (each $n = 21$: **effect +** /**effort −** , **effect +** /**effort +** , **effect −** /**effort −** , **effect −** /**effort +**). The procedure was the same as in the previous experiments. After the general task instructions page, each subject received one of the four versions of the problem. The text of all four problems is presented in Fig. 8. The four cards were: 6, 7, A, E, in all versions. The final instructions were: "Your task is to indicate which cards need to be turned over in order to establish whether what Mr. Bianchi said is true or false, at least as far as these four cards are concerned. Indicate only the cards that it is absolutely necessary to turn over".

11. Results

The frequencies of selection patterns in the four conditions are presented in Table 4. A factorial rank sum analysis, Effect (2) × Effort (2), for frequency tables (Meddis, 1984) was performed on data concerning correct versus incorrect selections. Consistent with our predictions, a main effect was found for effect, as well as for effort (in both cases, $z = 2.5$, $p < .01$). The two factors, effect and effort did not significantly interact. Pairwise comparisons (specific tests with procedure of correction for ties) indicated that, despite the presence of an abstract material, subjects produced a good rate of correct solutions (57%) in the **effect +** /**effort −** condition in which both relevance factors were predicted to affect performance in a positive way. This success rate did not significantly differ from that obtained in the two conditions in which only one relevance factor was predicted to improve performance (38% of correct solutions for both **effect +** /**effort +** and **effect −** /**effort −** conditions). In contrast, success rate in all these conditions was significantly superior to that obtained in the condition in which no

Table 4
Frequencies of selection patterns in the four conditions (Experiment 4, machine problem)

Pattern	Condition			
	Effect + /effort − (i.e., relevance)	Effect + /effort +	Effect − /effort −	Effect − /effort + (i.e., irrelevance)
P, not-Q	12	8	8	1
P, Q	1	5	5	9
P	2	3	0	5
not-P, Q	3	0	2	3
P, Q, not-Q	3	0	0	0
P, not-P	0	1	2	0
Q, not-Q	0	1	1	1
not-P, not-Q	0	1	2	0
All	0	0	0	2
Q	0	1	0	0
Not-P	0	0	1	0
None	0	1	0	0

$n = 21$ in all conditions.

relevance factor was supposed to affect performance (5% of correct solutions in the **effect − /effort +** condition; **effect + /effort −** vs. **effect − / effort +**: $z' = 3.6$, $p < .001$; **effect + /effort +** and **effect − /effort −** vs. **effect − /effort +**: $z' = 2.6$, $p > .01$).

Our predictions were confirmed. Both factors of relevance, effect and effort, were shown to play a role in performance.

GENERAL DISCUSSION

As shown by the results of our four experiments, the recipe suggested by Relevance Theory proved effective for constructing easy versions of the Selection Task. All predictions derived from Relevance Theory were confirmed. We want to argue that this account also explains past results, and does so in an integrated manner. We have already explained why the Selection Task should generally be difficult, and why the *P* card selection and the *P* and *Q* cards selection should be the most frequently encountered ones. We will focus here on the (correct) *P* and *not-Q* cards selections obtained with negative-consequent rules on the one hand, and with deontic tasks on the other.

12. Negative-consequent rules

Abstract Selection Tasks using rules with a negative consequent (e.g., "if a card has an A on the front, it does *not* have a 7 on the back") elicit a high rate (around 60%) of correct responses (for a review, see Evans *et al.*,

1993). Since, in this case, the rule is of the form "if P, then not Q", the correct selection is that of the *P* and *Q* cards. The *P* and *Q* cards selection is also the most frequent one in standard abstract versions, although, of course, here it is incorrect. It is tempting to think, then, that subjects are just indifferent to the presence of the negation in the negative-consequent versions. They choose, in both versions, the two cards mentioned in the rule (e.g., "A" and "7"), yielding a correct response in the negative-consequent versions, but for reasons that may have nothing to do with sound inference.

Evans (1984, 1989) explains the high rate of correct selections in negative-consequent tasks as the result of slightly more sophisticated but equally superficial processes of information selection. He invokes two pre-attentional heuristics. One heuristic, specific to conditionals, is cued by the word "if" and focuses on the conditional's antecedent. The other heuristic is more general; it focuses attention on the topic of any utterance. The topic is the same whether or not the main verb is negated. Given these heuristics, "if P then not Q" rules focus attention on P and on Q. Subjects select the corresponding *P* and *Q* cards, and, in the negative consequent case, reach the correct selection without any reasoning taking place.

Mental model theory (Johnson-Laird & Byrne, 1991) provides an account along similar lines. Subjects represent "if P then not Q" rules with two models: one where only the antecedent P is represented, and another where only the complementary Q of the consequent not-Q is represented, every negation yielding an explicit representation of what is negated. Since what is represented is what is selected, subjects select the *P* and *Q* cards.

Our relevance-theoretic account of the Selection Task suggests an alternative explanation of subjects' performance on negative-consequent versions. Unlike Evans (1984, 1989), we assume that, in all versions of the task, subjects go beyond superficial features of the rule. They envisage testing the rule through those of its logically derivable, directly testable consequences that they spontaneously infer. As indicated (see Introduction), the three consequences most pertinent to understanding subjects' performance are:

(a) The rule implies, of any given card having the feature *P*, that it has the feature *Q*.
(b) The rule, together with a presumption of relevance, implies in most contexts that there are cases of *P-and-Q*.
(c) The rule contradicts the assumption that there are cases of *P-and-(not-Q)*.

We pointed out that consequence **(c)** is much less easily inferable than consequence **(b)**, because **(c)** contains two negations: an implicit one in the attitude of contradiction, and an explicit one in the negation of *Q*.

However, when the rule is of the form "if P, then not-Q" **(a)**, **(b)**, and **(c)** become **(a′)**, **(b′)**, and **(c′)**:

(**a'**) The rule implies, of any given card having the feature *P*, that it has the feature *not-Q*.
(**b'**) The rule, together with a presumption of relevance, implies in most contexts that there are cases of *P-and-(not-Q)*.
(**c'**) The rule contradicts the assumption that there are cases of *P-and-Q*.

Now, deriving either consequence (**b'**) or consequence (**c'**) involves representing a negation: an explicit negation in the content of the implication in the case of (**b'**), and an implicit negation in the attitude of contradiction in the case of (**c'**). The direct advantage in terms of effort that (**b**) had over (**c**) is lost in the case of (**b'**) and (**c'**). Since representing explicit negation is more effort demanding than representing implicit negation (Horn, 1989), we hypothesize that the contradiction in (**c'**) is, if anything, easier to derive than the implicit negation in (**b'**).

On the effect side, it is again a point of general agreement that, in Strawson's terms, "the standard and primary use of 'not' is specifically to contradict or correct; to cancel a suggestion of oneself or another" (Strawson, 1952, p. 8; see also Horn, 1989). This is the case when the negation is used in the main clause, as for instance in a general conditional with a negative consequent. A statement of form [$\forall x$ ($Px \rightarrow$ not-Qx)] contradicts an assumption of form [$\exists x$ (Px and Qx)]. Thus (29) would typically be used to deny (27), or to deny a statement implying (27) such as (28):

(27) Some Parmesan cheese is smelly
(28) The Parmesan cheese I bought this morning is very smelly
(29) If it is Parmesan cheese, then it is not smelly

When an "if P, then not Q" rule is used in a thematic task, the content and context may indicate what assumption is being denied by the rule, or on the contrary may suggest that the rule is not to be taken as a denial (which will then cause poor performance as in Manktelow & Evans, 1979). When an "if P, then not Q" rule is used in an abstract Selection Task, without any prior assumption in the context standing to be denied, the artificiality of the task is even greater than usual. Still, we assume that subjects are disposed to, and capable of, constructing a pragmatically plausible interpretation of the rule. For this, they reconstruct the assumption that the rule should be taken to deny, that is, an assumption of form [$\exists x$ (Px and Qx)] (there are cases of *P-and-Q*). This interpretation of the rule as a denial causes them to make the correct selection, as explained and illustrated throughout this article. In other terms, the relevance-theoretic account of the Selection Task, together with independently well-motivated assumptions about the psycholinguistics of negation, predicts the very results found in negative-consequent versions of the task.

Applying Sperber and Wilson's Relevance Theory to the Selection Task

offers an alternative to, or an improvement of, Evans' own "Relevance Theory of the Selection Task" (Evans, 1989, 1993) that had been uniquely effective, up to now, in explaining the negative-consequent versions of the task. Though both views appeal to a notion of relevance, and it is conceivable that they could be made to cohere, we want at this stage to articulate the differences.

The gist of our difference with Evans regarding the cognitive processes that give rise to intuitions of relevance is the following. Evans distinguishes two types of thought processes which he calls *heuristic* and *analytic*. The function of heuristic processes is to select information relevant to the problem at hand. Heuristic processes are pre-attentive and driven by relatively superficial features of the stimulus, such as the linquistically determined topic of an utterance, or the focusing features of the word "if". Heuristic processes provide the input to analytic processes, where reasoning proper occurs. In the Selection Task, Evans argues, subjects' selections of cards "do not reflect *any* process of reasoning, in the sense of analytical processing, and are due entirely to heuristic processes" (Evans, 1984, p. 457). We, on the other hand, believe that human cognition is, at all levels of representation and processing, guided by expectations of relevance. These expectations may indeed by raised by superficial features of the stimulus, but also by deeper features recognized in, or attributed to the stimulus at any stage of its processing. Thus we do not accept Evans' distinction between successive heuristic and analytic *processes*. We would argue rather that most, if not all complex inferential processes have, to use Evans's own terms, both a "heuristic" and an "analytic" component, the heuristic component guiding the analytic "on line", and getting feedback from the analytic component's progress.

The more specific difference between Evans' and our account of negative-consequent versions of the task has, precisely, to do with the depth of understanding that might be attributed to subjects who make the correct *P* and *Q* cards selection. Beside considerations of generality of explanation, there are empirical reasons to assume that subjects understand well, and pay attention to, the logical structure of "if P then not Q" conditional statements.

Using a new experimental format, Cara and Broadbent (1995) required subjects to solve two tasks: an abstract Selection Task, and a conditional reasoning task in which subjects had to indicate all states of affair compatible with the two premises of the four standard conditional syllogisms (*modus ponens*, *modus tollens*, denial of the antecedent, and affirmation of the consequent). In both the Selection Task and the syllogisms, standard "if P then Q" and negative-consequent "if P, then not Q" conditional statements were used. Cara and Broadbent confirmed Evans' finding that the majority of subjects, in negative-consequent Selection Tasks, make the correct selection (69% vs. 4% with the "if P, then Q" rule). They also observed, however, that 90% of the same subjects gave the correct answers

to all four syllogistic problems when the major premise was of form "if P, then not-Q", versus 17% when the major premise was of form "if P, then Q". At least part of the reason why subjects perform better with "if P, then not Q" syllogisms than with "if P then Q" ones turns out to be that the latter, but not the former, lend themselves (for pragmatic reasons much discussed since Geis and Zwicky, 1971) to a biconditional reinterpretation.

Cara and Broadbent's results strongly suggest that subjects have a logically correct understanding of the negative-consequent conditional rule. It is still conceivable, in principle, that the same subjects, who give evidence of this deeper understanding when they are presented with the syllogistic task, stay at, or revert to, a wholly superficial understanding of the rule when they are presented with the Selection Task. However, this supposition would raise a new puzzle in order to solve the older one raised by performance observed on the negative-consequent versions of the Selection Task. Given the alternative – and more general – solution provided by the relevance-theoretic account, this unparsimonious supposition is unnecessary.

Further evidence for our approach is provided, we believe, by some recent innovative experiments of Evans himself. Evans (1994) presented several versions of the task on a computer screen and asked subjects to point with the mouse at the cards they were considering selecting. Subjects spent much more time considering the cards they ended up actually choosing. Thinking aloud protocols confirmed that subjects think about the cards they select, and hardly at all about the cards they do not select. From the point of view of Evans' relevance theory of the task, the fact that subjects reason "analytically" about the cards they select is somewhat puzzling, since, as Evans himself argues, this thinking does not affect their choices, which are and remain entirely based on "heuristic" procedures. Why should heuristic and analytic procedures converge in all cases? One might have expected that analytic reasoning would in some cases reverse the heuristic preference. In particular, one would have expected the heuristic appeal of the *Q* card in standard versions of the task to be countered at the analytic level.

From our point of view, subject's selection is based not on a superficial heuristic but on inferential processes that identify some combination of features that, given the rule, can be expected to occur, or, as the case may be, not to occur. Subjects' thinking about a card that they are about to select can be seen as their checking mentally that this card is indeed a potential instance of this combination of features. For instance, subjects may have inferred from the rule that there should be cases of *P-and-Q*. This focuses attentional processes on the *P* and on the *Q* cards. Subjects then satisfy themselves mentally that, when taken together, the hidden and the visible feature of each of these two cards might instantiate the combination *P-and-Q*. In Evans' account, the mutual consistency of non-inferential heuristic processes with analytic reasoning processes is puzzling. In our account, on the other hand, the mutual consistency of two causally and

semantically related processes, both inferential, is just what one would expect of a cognitively sound system.

13. Relevance in deontic versions of the task

Apart from the negative-consequent versions just discussed, the easiest Selection Tasks reported up to now in the literature have all been deontic versions. What makes a version deontic is not the linguistic form of the rule, but its interpretation in context. Thus the rule "If a card has an A on the front, it *must* have a 7 on the back" is commonsensically not interpreted as deontic, in spite of the modal verb "must" (which can always be understood as expressing epistemic or metaphysical necessity, anyhow). On the other hand, the rule "If a person is drinking beer, then the person is 19 years of age", though not containing a deontic "must", is easily interpreted deontically, that is, as stating not what people are actually doing, but what they should be doing. Purported deontic rules have to make sense. They must apply to entities which fall under deontic rules, that is, people, behaviors, or outcome of behaviors (Cheng & Holyoak, 1985; Cosmides, 1989; Gigerenzer & Hug, 1992; Girotto et al., 1989; Manktelow & Over, 1990, 1991).

The comparative ease with which subjects perform deontic tasks, suggests that a domain-specific competence is evoked. This idea has been developed in two different ways. Cheng and Holyoak (1985, 1989; see also Cheng et al., 1986) have argued that deontic versions of the task evoke acquired permission or obligation "pragmatic schema" at a level of abstraction intermediate between general domain-independent logical rules and knowledge of specific instances.[13] Cosmides (1989) has argued that the domain-specific competence evoked by deontic versions of the Selection Task is a phylogenetically determined "Darwinian algorithm", specialized for handling social contract situations, and useful, in particular, for detecting cheaters (see also Gigerenzer & Hug, 1992).

The two accounts differ not only with respect to the ontogenetic or phylogenetic origins of the mechanism, but also with respect to the manner in which the mechanism resolves the Selection Task. According to the pragmatic schema account, understanding a permission or an obligation involves grasping four individual rules, each of which affords a correct inference for one of the four cards. According to the Darwinian algorithm account, a look-for-cheater strategy focuses attention directly on the two

[13] Cheng and Holyoak use the phrase "domain-specific" in a non-standard way, to refer to the a view they oppose: the view that successful performance in the Selection Task is based on memory of specific instances. Their own view is nevertheless "domain-specific" in the usual sense (see Hirschfeld & Gelman, 1994).

cards representing potential cheaters. These two hypotheses, and the competition between them, have fostered a wealth of interesting research. Most of this research takes for granted the assumption that the difference in performance between the descriptive and the deontic versions is caused by domain-specific competences. This assumption itself develops from the premise that, if domain-general logical abilities determined performance, then one should expect no "content effects", and in particular no differences between the descriptive and the deontic versions of the task. This in turn presupposes that all versions of the task are logically equivalent and differ only by logically irrelevant aspects of their content. However this is blatantly not so.

Easy deontic versions are characterized not only by the fact that they contain a clear and sensible deontic rule, but also by the fact that what is at stake is not whether this rule is true or false (i.e., in force or not), but whether it is obeyed or not. While in the descriptive versions, subjects are asked to reason *about* the rule, in true deontic versions they are asked to reason *from* a rule given as axiomatic. When this is not the case and subjects are instructed to select evidence aimed at finding out – inductively – whether or not a deontic rule is actually in force, then their performance is poor and comparable to that of most descriptive versions (Cosmides, 1989, Experiment 6; Gigerenzer & Hug, 1992; Noveck & O'Brien, in press).

Given that the descriptive and the deontic tasks are logically different, it is not unreasonable to argue that, in spite of a superficial isomorphism, these are simply two different tasks (see Manktelow & Over, 1990, 1991, 1992). If so, then the difference in performance on the two tasks need not be seen as a genuine puzzle, and even less as a deep one. Subjects' good performance on the deontic task may be explained by assuming that people have good deontic reasoning abilities that are evoked by the deontic task. The difference in performance between the descriptive and the deontic tasks may be explained by the logical difference between the tasks, and not any more by logically unsound content effects. Deontic reasoning is, in a trivial sense, domain-specific – it is specific to the deontic domain – but it is much more general than either specific pragmatic schemas or specific Darwinian algorithms.

The issue is, however, opened anew by the evidence and the arguments we have presented here. We have shown that easy *descriptive* versions of the task could be devised at will. The contrast between difficult descriptive versions and easy deontic ones collapses. We have argued that, from a psychological point of view, it is less relevant to classify a task according to its formal logical structure, than according to the cognitive abilities it evokes. We have argued that the descriptive task evokes inferential comprehension mechanisms that determine intuitions of relevance, which in turn determine subjects' selections. When these intuitions go the right way (with respect to the task's requirements), selections are correct, and not otherwise. We want to suggest now that the same holds for the deontic task. If so,

in spite of their logical differences, deontic and descriptive versions are not psychologically so different after all.

We have argued, more specifically, that subjects succeed at the task when they interpret the descriptive conditional rule as a denial of the existence of *P-and-(not-Q)* cases. Similarly, we want to suggest that a deontic rule yields good performance when it is interpreted as forbidding the occurrence of *P-and-(not-Q)* cases. Factors parallel to those that favor such an interpretation in the descriptive case favor it in the deontic case. A deontic version, we predict, will yield good performance when *P-and-(not-Q)* cases are at least as easy to represent as *P-and-Q* cases, and when more effects are expected from knowing whether or not there are *P-and-(not-Q)* cases than from knowing whether or not there are *P-and-Q* cases.

Let us consider two classical examples in this light:

The drinking age problem: "If a person is drinking beer, then the person must be over 19 years of age" (Griggs & Cox, 1982; Fig. 2 above). *P-and-Q* is an individual drinking beer, and over 19 years of age. *P-and-(not-Q)* is an individual drinking beer and under 19 years of age. On the effort side, both are equally easy to represent. On the effect side, the drinking of beer by adults is a trivial event from which nothing significant follows. The drinking beer by young people tends to be seen as a social or moral problem and is therefore more significant, even in the absence of a legal prohibition. Given a rule in force against such under-age drinking, a violation of this rule implies moreover the possibility of sanctions. When subjects are instructed to play the part of a law-enforcer, then the presence of beer drinkers below 19 years of age implies that they should act, while the presence of lawful adult beer drinkers carries no particular implication at all.

The cholera problem: "If the form says 'ENTERING' on one side, the other side includes cholera among the list of diseases" (Cheng & Holyoak, 1985). Subject are instructed to play the role of an immigration officer at Manila airport checking passengers' inoculations. The cards shown are 'ENTERING' (*P* card), 'TRANSIT' (*not-P* card), 'cholera, typhoid, hepatitis' (*Q* card), 'typhoid, hepatitis' (*not-Q* card). On the effort side, representing a passenger entering the country and inoculated against a list of diseases including cholera (*P-and-Q*) is not easier than representing a passenger entering the country and inoculated against a list of diseases excluding cholera (*P-and-(not-Q)*), especially when the lists are short, and differ only by the presence of the word "cholera", so that the *not-Q* list is in fact shorter. On the effect side, as before, the presence of a *P-and-(not-Q)* case implies that steps must be taken by the immigration officer. The presence of a *P-and-Q* case has no particular consequences.

These two examples are typical of deontic versions that have elicited good performance. *P* and *not-P* are generally either complementary features (such as "ENTERING" vs. "TRANSIT"), or marked presence versus

unmarked absence of a feature (such as "alcoholic drink" vs. "non-alcoholic drink"). *Q* and *not-Q* are also generally either complementary features (such as "over 19" vs. "under 19"), or marked presence versus unmarked absence of a feature (such as "list with 'cholera'" vs. "list without 'cholera'").[14] Generally, neither *P* nor *not-P* combines with either *Q* or *not-Q* to form a standard concept (marked, say, by lexicalization). With most deontic rules, therefore, all four combinations – *P-and-Q*, *(not-P)-and-Q*, *P-and-(not-Q)*, and *(not-P)-and-(not-Q)* – seem at first sight to be roughly as easy to represent.

In fact, in the case of many deontic rules, the *P-and-(not-Q)* case is linguistically highlighted and is easier to represent than all the other. There is a wealth of terms to designate violators and violations of specific rules or types of rules, for example, adultery, arson, blasphemy, embezzlement, murder, plagiarism, rape, swindle, treason. There is no correspondingly rich terminology to designate non-adultery, non-arson, non-blasphemy, etc. While the specific rules that have been used in deontic Selection Tasks do not carry specific names for their violations, all these violations come at least under the label "violation", or even under more specific labels such as "cheating" or "fraud". Again, there are no counterpart one-word labels for the many varieties of rule-abiding, or even for rule-abiding in general. *P-and-(not-Q)* events are instances of at least one lexicalized concept available to subjects: that of violation, whereas, in general, *P-and-Q* events don't belong to a distinctive named category. On the effort side, therefore, we expect *P-and-(not-Q)* violations of deontic rules to be at least as easy, and probably easier, to represent than *P-and-Q* cases of conformity.

We assume that subjects understand the task. That is, they understand that they are asked whether the rule is being obeyed or disobeyed. They understand that conformity to a conditional rule is not directly testable. They understand therefore that they must infer from the rule directly observable states of affairs allowed or forbidden by the rule. The situation so far is parallel to the descriptive case. We might be tempted to envisage that subjects would draw from a deontic rules consequences such as **(a″)**, **(b″)**, and **(c″)** which are parallel to the **(a)**, **(b)**, and **(c)** consequences of the descriptive versions:

(a″) The rule requires, of any given item having the feature *P*, that it have the feature *Q*.
(b″) ? The rule, together with a presumption of relevance, requires in most contexts that there be cases of *P-and-Q*.
(c″) The rule forbids that there be cases of *P-and-(not-Q)*.

[14] *A contrario*, Girotto et al. (1992) showed that, in the case of abstract deontic rules, when *not-Q* is not explicitly defined as the complementary of *Q*, subjects tend to select only the *P* card (see also Griggs & Cox, 1993; Jackson & Griggs, 1990).

However, while (**a**″) and (**c**″) are unproblematically correct inferences, the (**b**″) case differs significantly in this respect from its descriptive (**b**) counterpart. Conditional deontic rules are generally not instituted in order to bring about the occurrence of *P-and-Q* cases (e.g., adult beer-drinkers). Rather, given that there are, or that there might be *P* cases (e.g., beer-drinkers), conditional deontic rules are instituted in order to ensure that these *P* cases should also have the feature *Q* (e.g., adulthood).

The fact that the rule expresses an enforced regulation or an accepted contract means, in most situations, that the occurrence of violations is much more relevant than that of permissible behavior. When, on top of that, subjects are instructed to play the role of a law enforcer, or of someone who stands to suffer from violations, then expectations of effect are wholly on the *P-and-(not-Q)* side. *A contrario*, when the rule seems wholly arbitrary to subjects, and is presented in such a manner that they don't see what effects violations would carry, the rate of correct answers falls (Girotto et al., 1989, Experiment 3).

Of particular interest, from a Relevance Theory point of view, is the case of perspective shifts studied by Gigerenzer and Hug (1992), Manktelow and Over (1991), and Politzer and Nguyen-Xuan (1992) (see also Light et al., 1990). Some deontic rules express a contract relation between two parties such that the state of affairs described in the antecedent can be brought about by one party to the contract, while the state of affairs described in the consequent can be brought about by the other party. Each party can benefit the other party at a cost to itself, and the contract makes for each party getting the benefit contingent on paying the cost.

Take as an illustration the day off experiment of Gigerenzer and Hug (1992). The rule was:

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

When subjects were cued in the perspective of an employee, 75% made the correct selection of the *P* and *not-Q* cards. However, when they were cued in the perspective of the employer, 61% of Gigerenzer and Hug's subjects chose the *not-P* and the *Q* cards, a pattern of selection rarely reported for descriptive versions of the task,[15] and a logically incorrect one (on the assumption that subjects are considering checking whether the explicit rule is violated).

The explanations that have been given for subjects' performance with such perspective shifts have all been relatively domain-specific. Going from the most to the least domain-specific proposal:

[15] This pattern of response was first reported by Cosmides (1989) with "switched" deontic rules. However these rules were pragmatically awkward in the context, unlike the rules used in the experiments we are discussing in this section (see also Mosconi & D'Urso, 1974).

- Gigerenzer and Hug (1992), drawing on Cosmides (1989), argue that a domain-specific "look-for-cheater" strategy determines different selections depending on the perspective: the *not-P* and *Q* selection is the appropriate one to identify cheating employees from the perspective of the employer.
- Politzer and Nguyen-Xuan (1992) argue that the two perspectives yield two different interpretations of the rule, each activating a different "pragmatic schema" (in the sense of Cheng & Holyoak, 1985).
- Manktelow and Over (1991) argue that subjects' performance is guided by a calculus of utilities that affects reasoning on deontic tasks (but not on descriptive tasks).

For Gigerenzer and Hug (1992) and Manktelow and Over (1991), subjects' responses when the perspective is reversed are not correct solutions of the task construed in a narrowly logical way, but they are nevertheless rational solutions, based on sound domain-specific mechanisms. For Politzer and Ngyuen-Xuan (1992), when perspective is reversed, the rule is interpreted differently, and subjects provide a logically correct solution to the problem as contextually reinterpreted. Politzer and Ngyuen-Xuan fail to consider the possibility that the linguistic pragmatic factors they rightly see as determining the interpretation of the problem, may also determine subjects' intuitions of relevance and ultimately their responses, without any appeal to "pragmatic schemas" or any other inferential competence specific to the deontic sphere.

From a linguistic pragmatic point of view, a simple conditional rule makes explicit the contractual rights and duties of only one of the parties, and leaves the rights and duties of the other party implicit. For instance, in the day off problem, the other, implicit half of the contract could be made explicit as:

(31) If an employee gets a day off during the week, then that person must work on the weekend

More generally, a rule of the form (32) understood as a partial representation of reciprocal contract may be taken to convey an implicit converse rule of the form (33):

(32) $[\forall x\ (Px \rightarrow \text{must}\ Qx)]$
(33) $[\forall x\ (Qx \rightarrow \text{must}\ Px)]$

When will an implicit content be retrieved? The relevance-theoretic answer is: when retrieving this implicit content is a step on the least-effort path to achieving the expected level of relevance. By instructing subjects to adopt the perspective of the party whose right and duties are left implicit – for instance, the perspective of the employer in the day off experiment –

and by providing adequate information, the experimenter causes subjects to form expectations of relevance that can only be satisfied by retrieving, to begin with, this implicit content. This implicit content itself contributes to relevance through *its* consequences, and it is on the bases of these that subjects are then predicted to answer the Selection Task. These consequences are of the form:

(**a'''**) The rule requires, of any given item having the feature Q, that it have the feature P.
(**c'''**) The rule forbids that there be cases of *Q-and-(not-P)*.

In the day off problem, for instance, subjects instructed to take the perspective of the employer should select the Q card because of both (**a'''**) and (**c'''**), and the *not-P* card because of (**c'''**). From the relevance-theoretic point of view, then, subjects are seen as, first, using sound comprehension procedure to retrieve the implicit content of the rule, and, then, as applying logically sound, domain-general inferential procedures to this implicit content.

All the available evidence from studies on deontic versions of the Selection Task, including the recent evidence coming from perspective shift, is predicted by our wholly general relevance-theoretic account. So, this account should be preferred on grounds of generality. Does this mean that there are no domain-specific competences in the deontic domain? No, it does not. There may well be independent reasons to assume that there exist such competences. Cosmides (1989) is right in arguing that evolution would have favored individuals better able to detect cheaters, and Cheng and Holyoak (1985) are right in drawing attention to the intense learning of rules of permission and obligation in the socialization process. We pointed out the existence of a much richer vocabulary for violations than for cases of conformity; this too speaks for the existence of a domain-specific cognitive competence in the area. The Selection Task, however, is not a good tool to study these domain-specific mechanisms, because relevance-guided comprehension procedures that are anyhow at work are capable of producing on their own the same intuitions of relevance that these mechanisms are supposed to explain.

There are three possibilities, then, among which the available Selection Task evidence cannot decide. The first possibility is that there are no mechanisms specific to the deontic domain (or part thereof) at all. The second possibility is that comprehension mechanisms just short-circuit domain-specific mechanisms. The third possibility is that the comprehension mechanisms and domain-specific mechanisms jointly contribute to subjects' performance, but in this case their effects are, as things stand, confounded. We see this third possibility as plausible. Should one then try to separate the effects of the two kind of factors in the Selection Task? If one's interest is not in the psychologists' toolkit *per se*, but in the psychologists' subject

matter, that is, the human mind, then it would seem more productive to look for other tools to study domain-specific competences in the deontic sphere.

14. Conclusion

We have argued that most subjects' selection in all true versions of the Selection Task result from the following procedure. Subjects infer from the rule directly testable consequences. They infer them in their order of accessibility, and stop when the resulting interpretation of the rule meets their expectations of relevance. Both the order of accessibility of the consequences and expectations of relevance may vary with the content and context of the rule. Subjects then select the cards that may test the directly testable consequences they have inferred from the rule. We have predicted that, in order to cause subjects to select, correctly, the *P* and the *not-Q* cards, one should make the consequence of the rule [there are no *P-and-(not-Q)* cases] at least as accessible as, and richer in cognitive effects than, the consequence [there are *P-and-Q* cases]. We have provided a detailed recipe for doing just that, and we have confirmed our prediction with four experiments. We have argued that past results properly reanalyzed confirm this account.

Relevance Theory has been initially developed on the basis of philosophical arguments, general psychological considerations, and empirical work in linguistics. There has been little experimental work testing the theory, and only relatively peripheral implications of the theory have been tested (see Foster-Cohen, 1994; Jorgensen, Miller, & Sperber, 1984; Happé, 1993; Watson, in press). The present study tests and confirms core assumptions of Relevance Theory. It also illustrates how precise and testable explanatory hypotheses can be derived from the theory.

If it is properly explained by Relevance Theory, then the Selection Task highlights the difference between spontaneous inference and reflective reasoning. It does not, however, reveal anything profound about reasoning proper. It does not provide evidence for or against different theories of human reasoning such as mental logic (Braine & O'Brien, 1991; Braine & Rumain, 1983; Rips, 1994), mental models (Johnson-Laird, 1983; Johnson-Laird & Byrne 1991), pragmatic schemas (Cheng & Holyoak, 1985; Cheng et al., 1986), and Darwinian algorithms (Cosmides, 1989). In the Selection Task, the various reasoning processes that have been suggested on the basis of these theories are pre-empted by processes that determine intuitions of relevance. This has long been argued by Evans (1984, 1989). Unlike Evans, however, for whom relevance is identified differently in different versions of the task, we see exactly the same procedure at work in all versions, whether descriptive or deontic. Unlike Evans, also, we see this procedure as one of inferential comprehension.

We did attribute to our subject a capacity to perform spontaneously specific deductive inferences involving quantifiers, and a capacity to recognize specific contradictions. This raises the following question: what are the capacities that make these inferences possible? Selection Task data, however, do not help answer this question. Such capacities can be as well described in terms of mental logic as of mental models, and these do not exhaust all the possibilities. Here again, then, the Selection Task does not help discriminate among theories of reasoning.

But what if one wants to study precisely the inferential capacities at work in comprehension, the existence of which is well in evidence in the Selection Task? Even then, the Selection Task itself is a tool of limited interest, because the puzzling results it has brought to light are, if our analysis is correct, largely artifactual. In most versions of the task where subjects fail to make the logically correct selection, the occurrence of the feature Q is not inferable from that of the feature P, and the use of the conditional is therefore pragmatically infelicitous – and often the whole problem is pragmatically infelicitous in more ways than one. Comprehension procedures are geared towards the processing of optimally relevant communicative behaviors (Sperber & Wilson, 1986), and they don't work as well as stimuli that not only depart from optimal relevance, but moreover do so in arbitrary ways. When such stimuli are used, subjects' responses are generally underdetermined by the structure of the task, as is shown by the fact that these responses tend to be scattered among many different patterns. With an adequate pragmatic approach, versions of the task can be devised where most subjects' intuitions of relevance go the right way. This is of some interest for the study of the pragmatics of conditionals. It would seem to be of marginal value only for the experimental study of inferential comprehension in general.

If there is one very general positive lesson to be learnt from this work on the Selection Task, it is that people are nearly-incorrigible "cognitive optimists". They take for granted that their spontaneous cognitive processes are highly reliable, and that the output of these processes does not need re-checking. Just as they trust their perceptions, they trust their spontaneous inferences and their intuitions of relevance. In most situations, they may well be right. Still, all one has to do to make people appear irrational is to devise an experiment where their intuitions of relevance go the wrong way.[16] Psychologists too may be subject to illusions of relevance. The strong sense among them that Selection Task results are highly relevant to our understanding of the human mind may have been, we submit, precisely such an illusion.

[16] See Politzer (1993) for a detailed discussion of many standard reasoning experiments from a similar point of view.

Acknowledgements

This research was made possible by grants from the Ministère de la Recherche, Paris, the Programme Cognisciences, CNRS, France, and the CNR, Roma. It benefited from the institutional support of the CREA, Ecole Polytechnique, Paris. We thank Daniel Andler, Frank Döring, Denis Hilton, Philip Johnson-Laird, Ira Noveck, Mike Oaksford, David Over, Guy Politzer, François Recanati, Rita Watson, Deirdre Wilson, and an anonymous reviewer for their comments and suggestions on earlier drafts of this article.

References

Anderson, J.R. (1986). *Cognitive psychology and its implications*. San Francisco: Freeman.

Bach, K., & Harnish, R. (1979). *Linguistic communication and speech acts*. Cambridge, MA: MIT Press.

Begg, I., & Harris, G. (1982). On the interpretation of syllogisms. *Journal of Verbal Learning and Verbal Behavior, 21*, 595–620.

Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.

Blakemore, D. (1992). *Understanding utterances*. Oxford: Blackwell.

Blass, R. (1990). *Relevance relations in discourse: a study with special reference to Sissala*. Cambridge, UK: Cambridge University Press.

Braine, M.D.S., & O'Brien, D.P.O. (1991). A theory of *if*: a lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98*, 198–203.

Braine, M.D.S., Reiser, B.J., & Rumain, B. (1984). Some empirical justifications for a theory of natural propositional logic. In G.H. Bower (Ed.), *The psychology of learning and motivation*, (Vol. 18). New York: Academic Press.

Braine, M.D.S., & Rumain, B. (1983). Logical reasoning. In J.H. Flavell & E.M. Markman (Eds.), *Handbook of child psychology, Vol. 3: cognitive development* (4th ed.). New York: Wiley.

Cara, F., & Broadbent, S. (1995). Conditional inferences and card selections in Wason's task. Submitted manuscript.

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

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

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

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

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

D'Andrade, R.G. (1989). Culturally based reasoning. In A. Gellatly, D. Rogers, & J.A. Sloboda (Eds.), *Cognition and social worlds*. Oxford: Clarendon Press.

Evans, J.St.B.T. (1972). Interpretation and "matching bias" in a reasoning task. *British Journal of Psychology, 24*, 193–199.

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

Evans, J.St.B.T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, *75*, 451–468.

Evans, J.St.B.T. (1989). *Bias in human reasoning: causes and consequences*. Hillsdale, NJ: Erlbaum.

Evans, J.St.B.T. (1993). Bias and rationality. In K.I. Manktelow & D. Over (Eds.), *Rationality*. London: Routledge.

Evans, J.St.B.T. (1994). Relevance and reasoning. In S. Newstead & J.St.B.T. Evans (Eds.), *Current directions in thinking and reasoning*. Hillsdale, NJ: Erlbaum.

Evans, J.St.B.T., & Lynch, J.S. (1973). Matching bias in the selection task. *British Journal of Psychology*, *64*, 391–397.

Evans, J.St.B.T., Newstead, S.E., & Byrne, R.M.J. (1993). *Human reasoning: the psychology of deduction*. Hillsdale, NJ. Erlbaum.

Evans, J.St.B.T., Over, D., & Manktelow, K. (1993). Reasoning, decision making and rationality. *Cognition*, *49*, 165–187.

Fiedler, K., & Hertel, G. (1994). Content-related schemata versus verbal-framing effects in deductive reasoning. *Social Cognition*, *2*, 129–147.

Foster-Cohen, S.H. (1994). Exploring the boundary between syntax and pragmatics: relevance and the binding of pronouns. *Journal of Child Language*, *21*, 237–255.

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

Gigerenzer, G., & Hug, K. (1992). Reasoning about social contracts: cheating and perspective change. *Cognition*, *43*, 127–171.

Girotto, V., Blaye, A., & Farioli, F. (1989). A reason to reason: pragmatic analysis of children's reasoning. *European Bulletin of Cognitive Psychology*, *9*, 297–321.

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

Girotto, V., Mazzocco, A., & Cherubini, P. (1992). Judgements of deontic relevance in reasoning: a reply to Jackson and Griggs. *Quarterly Journal of Experimental Psychology*, *44A*, 469–482.

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

Griggs, R.A. (1989). To "see" or not to "see": that is the selection task. *Quarterly Journal of Experimental Psychology*, *41A*, 517–529.

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. (1993). Permission schemas and the selection task. *Quarterly Journal of Experimental Psychology*, *46A*, 637–651.

Griggs, R.A., & Jackson, S.L. (1990). Instructional effects on responses in Wason's Selection Task. *British Journal of Psychology*, *81*, 197–204.

Happé, F.G.E. (1993). Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition*, *48*, 101–119.

Hempel, C.G. (1965). *Aspects of scientific explanation*. New York: Free Press.

Hirschfeld, L.A., & Gelman, S.A. (1994). *Mapping the mind: domain specificity in cognition and culture*. Cambridge, UK: Cambridge University Press.

Hoch, S.J., & Tschirgi, J.E. (1985). Logical knowledge and cue redundancy in deductive reasoning. *Memory and Cognition*, *13*, 453–462.

Horn, L.R. (1989). *A natural history of negation*. Chicago: Chicago University Press.

Ifantidou-Trouki, E. (1993). Sentential advers and relevance. *Lingua*, *90*, 69–90.

Jackson, S.L., & Griggs, R.A. (1988). Education and the selection task. *Bulletin of the Psychonomic Society*, *26*, 327–330.

Jackson, S.L., & Griggs, R.A. (1990). The elusive pragmatic schemas effect. *Quarterly Journal of Experimental Psychology*, *42*(A), 353–373.

Johnson-Laird, P.N. (1983). *Mental models*. Cambridge, UK: Cambridge University Press.

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

Johnson-Laird, P.N., & Byrne, R.M.J. (1992). Modal reasoning, models and Manktelow & Over. *Cognition*, *43*, 173–182.

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

Johnson-Laird, P.N., & Wason, P.C. (1970a). Insight in a logical relation. *Quarterly Journal of Experimental Psychology*, *22*, 49–61.

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

Jorgensen, J., Miller, G.A., & Sperber, D. (1984). Test of the mention theory of irony. *Journal of Experimental Psychology: General*, *113*, 112–120.

Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

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

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, *94*, 211–228.

Leech, G. (1983). *Principles of pragmatics*. London: Longman.

Legrenzi, P. (1970). Relations between language and reasoning about deductive rules. In G.B. Flores D'Arcais & W.J.M. Levelt (Eds.), *Advances in psycholinguistics*. Amsterdam: North-Holland.

Legrenzi, P., Girotto, V., & Johnson-Laird, P.N. (1993). Focussing in reasoning and decision making. *Cognition*, *49*, 37–66.

Lehman, D.R., & Nisbett, R.E. (1990). A longitudinal study of the effects of undergraduates' training on reasoning. *Developmental Psychology*, *26*, 952–960.

Levinson, S. (1983). *Pragmatics*. Cambridge, UK: Cambridge University Press.

Liberman, N., & Klar, Y. (in press). Hypothesis testing in Wason's selection task: social exchange, cheating detection or task interpretation. *Cognition*.

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

Love, R., & Kessler, C. (in press). Focusing in Wason's selection task: content and instruction effects. *Thinking and Reasoning*.

Manktelow, K.I., & Evans, J.St.B.T. (1979). Facilitation of reasoning by realism: effect or not-effect? *British Journal of Psychology*, *70*, 477–488.

Manktelow, K.I., & Over, D.E. (1990). *Inference and understanding: a philosophical approach*. London: Routledge.

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

Margolis, H. (1987). *Pattern, thinking and cognition: a theory of judgement*. Chicago: University of Chicago Press.

Meddis, R. (1984). *Statistics using ranks: a unified approach*. Oxford: Blackwell.

Moeschler, J. (1989). Pragmatic connectives, argumentative coherence and relevance. *Argumentation*, *3*(3), 321–339.

Mosconi, G. (1990). *Discorso e pensiero* [Discourse and thinking]. Bologna: Il Mulino.

Mosconi, G., & D'Urso, V. (1974). *The selection task from the standpoint of the theory of double code*. Paper presented at the Conference on Selection Task, University of Trento, Italy.

Newstead, S.E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language*, *28*, 78–91.

Newstead, S.E., & Griggs, R.A. (1983). Drawing inferences from quantified statements: A study of the square of oppositions. *Journal of Verbal Learning and Verbal Behavior*, *22*, 535–546.

Noveck, I.A., & O'Brien, D.P. (in press). To what extent do pragmatic reasoning schemas affect performance on Wason's selection task? *Thinking and Reasoning*.

Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K.I. Manktelow & D. Over (Eds.), *Rationality*. London: Routledge.

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

O'Brien, D.P. (1994). Finding logic in human reasoning requires looking in the right places: a mental-logic view of logical-reasoning tasks. In S. Newstead & J.St.B.T. Evans (Eds.), *Current directions in thinking and reasoning*. Hillsdale, NJ: Erlbaum.

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

Politzer, G. (1990). Immediate deduction between quantified sentences. In K.J. Gilhooly, M.T. Keane, R.H. Logie, & G. Erdos (Eds.), *Lines of thinking: reflections on the psychology of reasoning*. Chichester: Wiley.

Politzer, G. (1993). *La psychologie du raisonnement: lois de la pragmatique et de la logique formelle*. Thèse de Doctorat d'Etat, Université de Paris VIII.

Politzer, G., & Nguyen-Xuan, A. (1992). Reasoning about promises and warnings: Darwinian algorithms, mental models, relevance judgements or pragmatic schemas? *Quarterly Journal of Experimental Psychology, 44A*, 401–421.

Pollard, P., & Evans, J.St.B.T. (1983). The effect of experimentally contrived experience on reasoning performance. *Psychological Research, 45*, 287–301.

Pollard, P., & Evans, J.St.B.T. (1987). Content and context effects in reasoning. *American Journal of Psychology, 100*, 41–60.

Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson.

Quine, W. (1960). *Word and object*. Cambridge, MA: Harvard university Press.

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

Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: non-consequential reasoning and choice. *Cognitive Psychology, 24*, 449–474.

Smith, E.E., Langston, C., & Nisbett, R.E. (1992). The case for rules in reasoning. *Cognitive Science, 16*, 1–40.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L.A. Hirschfeld & S.A. Gelman (eds.), *Mapping the mind: domain specificity in cognition and culture*. Cambridge, UK: Cambridge University Press.

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

Sperber, D., & Wilson, D. (in press). Postface. In *Relevance: communication and cognition* (2nd ed.). Oxford: Blackwell.

Stich, S. (1990). *The fragmentation of reason*. Cambridge, MA: Brandford Books/MIT Press.

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

Sweetser, E. (1990). *From etymology to pragmatics*. Cambridge, UK: Cambridge University Press.

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

Van Fraasen, B.C. (1980). *The scientific image*. Oxford: Clarendon Press.

Wason, P.C. (1959). The processing of negative and positive information. *Quarterly Journal of Experimental Psychology, 11*, 92–107.

Wason, P.C. (1966). Reasoning. In B. Foss (Ed.), *New Horizons in Psychology*. Harmondsworth: Penguin.

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

Wason, P.C. (1969). Regression in reasoning? *British Journal of Psychology, 60*, 471–480.

Wason, P.C. (1977). Self-contradictions. In P.N. Johnson-Larid & P.C. Wason (Eds.), *Thinking: readings in cognitive science*. Cambridge, UK: Cambridge University Press.

Wason, P.C., & Green, D. (1984). Reasoning and mental representation. *Quarterly Journal of Experimental Psychology*, *36A*, 597–610.

Wason, P.C., & Johnson-Laird, P.N. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, *61*, 509–515.

Wason, P.C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, *23*, 63–71.

Watson, R. (in press) Relevance and definition. *Journal of Child Language*.

Wilson, D., & Sperber, D. (1993). Linguistic form and relevance. *Lingua*, *90*, 1–25.

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

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