Rationality and the experimental study of reasoning

Guy Politzer

In P. Bourgine & J.-P. Nadal (Eds.) (2004). *Cognitive economics: an interdisciplinary approach* (pp. 79-93). Berlin: Springer-Verlag.

Rationality and the experimental study of reasoning

Guy Politzer

CNRS, Saint-Denis

Abstract. A survey of the results obtained during the past three decades in some of the most widely used tasks and paradigms in the experimental study of reasoning is presented. It is shown that, at first sight, human performance suffers from serious shortcomings. However, after the problems of communication between experimenter and subject are taken into account, which leads to clarify the subject's representation of the tasks, one observes a better performance, although still far from perfect. Current theories of reasoning, of which the two most prominent are very briefly outlined, agree in identifying the load in working memory as the main source of limitation in performance. Finally, a recent view on human rationality prompted by the foregoing results is described.

1. Introduction

Two areas of research in psychology are relevant to the study of human rationality. One is the study of reasoning and problem solving. A sample of the most important and typical laboratory tasks used in the study of reasoning together with an overview of performance on these tasks will be presented. It will be followed by a critical assessment of the results and of the interpretation commonly made of them by investigators in this field. Economists are at the origin of the second area of research, the study of judgment under uncertainty and decision making. Its main results are well known and we will limit ourselves to an examination of a few of them on the same methodological basis used to assess the results of the first area. Finally, it will be shown that the recent trend which brings together the two areas of research offers new insights into the nature of human rationality by suggesting the existence of two types of rationality.

2. Studies of reasoning in the laboratory

In a typical reasoning task, some preliminary instructions regarding the general procedure and, if appropriate, some rough explanations about the aim of the experiment are given to the participant, following which the information proper relevant to the task is provided. This generally consists of (i) the statement of either just the premises of an argument or the whole argument, often inserted in a scenario which provides a context; (ii) a request for completing the argument with a conclusion (production task) or for assessing the provided conclusion (evaluation task). Nowadays, the aim of such experiments is to test hypotheses derived from theory, or even to perform tests between rival theories; nevertheless, it is possible to exploit the results from a normative point of view, provided one can identify the logical argument treated by the participant, a critical requirement which cannot be taken for granted. As we will see later, this argument is not always the one intended by the experimenter.

The experimenter may interact with the participant in a variety of ways. Perhaps the commonest way to gather data is still through questionnaires generally administered in group. Presentation of the test materials and data collection may also be obtained individually by using a computer or orally in a face-to-face interview. Most individuals tested are University students who are seldom paid for their participation. In the next two sections, a number of classic tasks, deductive and inductive, will be reviewed.

2.1. Studies of deduction

Following standard logic, psychologists distinguish propositional and predicate reasoning.

2.1.1. Propositional reasoning.

Within this domain, the most thoroughly investigated task, to which we will limit ourselves, is called *conditional reasoning*. It is based on four arguments, two deductively valid:

- Modus Ponendo Ponens (MP): *if P then Q; P; therefore Q,* and
- Modus Tollendo Tollens (MT): *if P then Q; not-Q; therefore not-P,*

and two deductively invalid, which are the fallacies of:

- Affirming the Consequent (AC): *if P then Q; Q; therefore P,* and
- Denying the Antecedent (DA): *if P then Q; not-P; therefore not-Q.*

As with other kinds of argument, there are various ways to instantiate the sentences; using a scenario has the disadvantage that while evaluating the truth of the conclusion (as well as that of the premises) participants can be influenced by their knowledge or beliefs. This is why experimenters often create a rather decontextualised microworld, while also avoiding over-abstracting; this can be achieved, for instance, by asking participants to imagine a metallic board and magnetic letters that can be put on the board. With this material, the four arguments are as follows (referring to what there is on the board, and using arbitrary letters):

MP: if there is a T, there is a P; there is a T; therefore there is a P.

MT: if there is a D there is a G; there is not a G; therefore there is not a D.

AC: if there is a K there is a W; there is a W; therefore there is a K.

DA: if there is a V there is a B; there is not a V; therefore there is not a B.

Participants are required to decide whether or not the conclusion follows necessarily from the premises (or in a variant of the task in which no conclusion is provided, to indicate what, if anything, follows from the premises). Performance is robust. Considering the two valid arguments, nearly everybody endorses the conclusion of MP, but only about two thirds of the participants endorse that of MT. For the two invalid arguments, between one third and two thirds correctly answer that no conclusion follows (Evans, Newstead and Byrne, 1993). In summary, people hardly ever err on MP, one third err on MT, and about one half fall prey to the two fallacies of AC and DA. These experimental data seem to confirm and specify observations already made by the Greeks.

2.1.2. Predicate reasoning.

The two tasks that are the most studied in predicate reasoning date back to Aristotle. The first one is called *immediate inference;* it consists of elementary one-premise arguments in which the premise and the conclusion are standard quantified sentences, i. e., sentences of the type subject-predicate starting with *all, some,* or *no.* In experiments, participants are presented with one premise such as, e. g. *[in a bag of marbles] some white marbles are small*, and asked to evaluate (by 'true', 'false', or 'one cannot know') one or several conclusions suggested to them, such as *all white marbles are small; no white marble is small,* etc. *All…* and *no…* sentences are called *contraries; all…* and *some… not* sentences, as well as *no…* and *some…*sentences are called *contradictories.* Inference within a pair of contraries or a pair of contradictories logically yields a 'false' response, which is the answer given by most people. The picture differs sharply for the other inferences. Inferences from *all…* to *some…* and from *no…* to *some…not…* are valid and logically require a 'true' response, while in the other direction they are invalid and require a 'one cannot know' response. However, only about one quarter of the responses coincide with the formal logical response; a large majority opt for the response 'false' in both directions. Similarly, most people respond 'true' instead of the formal logical response 'one cannot know' to the conclusion *some…* drawn from the premise *some… not…* and vice versa (Begg and Harris, 1982; Newstead and Griggs, 1983; Politzer, 1990). In sum, performance seems to be very high on one part of these inferences but surprisingly low on the other part.

The other task studied in predicate reasoning is the most famous of all reasoning tasks, namely Aristotle's categorical syllogisms. These are two-premise arguments using the four standard quantified sentences; instead of involving four different classes (the subject and the predicate of the first premise and the subject and the predicate of the second premise) syllogisms involve only three classes because one class is common to either both subjects, or both predicates, or the subject of one premise and the predicate of the other. The common class is called the middle term and it never appears in the conclusion, so that the latter has the other two classes as subject and predicate. These constraints determine 64 possible pairs of premises out of which 27 yield a conclusion that follows validly from them. Here are two examples of syllogisms:

(At a party) (i) all the players are drinkers; no smoker is a drinker; what, if anything, follows ? (ii) all the drinkers are players; no smoker is a drinker; what, if anything, follows ? (the solution is at the end of this chapter). As the reader can see, the difficulty is extremely variable: most people get the first syllogism right whereas most fail the second one. The percentage of correct solution varies from 10% for the hardest syllogisms to 90% for the easiest; performance is higher on valid than invalid syllogisms (in other words, on the whole, it is easier to find the right conclusion of valid syllogisms than to recognise the absence of a conclusion for invalid syllogisms). A number of response biases have been described; a bias is a feature that characterises an erroneous answer in a systematic manner. In the present case, it was observed long ago that participants have a tendency to give a conclusion that reproduces the properties of the premises (e. g., if both premises are negative, give a negative conclusion, and so on.) Another bias consists in choosing as the subject or the predicate of the conclusion a class that is a subject or a predicate, respectively, in the premises; for instance, in the second syllogism above, the tendency is to choose a conclusion in which *smoker* is the subject (as in the second premise) and *player* the predicate (as in the first premise). There is also a belief bias: it consists of a tendency for people to accept the conclusions that coincide with their beliefs, and more markedly to reject the conclusions that contradict their beliefs, keeping constant logical validity (Evans, Newstead and Byrne, 1993). In brief, through these various biases - and to the extent that people are responsive to nonlogical surface features or content (as opposed to formal structure) - people seem to exhibit illogicality.

2.2. Studies of induction

Unlike deductively valid arguments, the conclusions of which are necessarily true if the premises are true, inductive arguments deliver a conclusion whose truth is more or less strongly plausible. One of the oldest inductive tasks studied in the laboratory is the *letter series completion task* (Simon and Kotovsky, 1963; Kotowski and Simon, 1973). Participants are presented with a series of letters and requested to indicate the missing letter that logically completes the series. Here are a few examples:

1) c d c d c d. .
2) a t b a ta a t b a t. .
3) m a b m b c m c d m. .
4) w x a x y b y z c z a d a b. .

These examples are presented here in increasing order of difficulty; hardly anybody fails to give the correct answer to the first problem, 10% fail the second, one quarter the third, and one third the last one.

Two tasks designed to study lay people's hypothesis testing behaviour have attracted most of psychologists' work. The first one is called the *2 4 6 task* (Wason, 1960). The situation is that of a game played between the experimenter and the participant. The former chooses a rule to generate sequences of three numbers and the latter has to discover this rule. In order to do so, there are two sources of information. One is the result of tests made by the participant: he chooses triples and submits them to the experimenter, who replies every time by 'yes' (the triple obeys the rule) or 'no' (it does not). (ii) The other source of information is provided by the experimenter at the beginning of the game; the participant is told that the triple *2, 4, 6* is an example that conforms to the rule. When the participants think they have discovered the rule, they state it; if they are wrong, the game continues for another cycle until the stated rule is correct or the participants give up. The rule followed by the experimenter is *three increasing numbers* (integers). It is usually observed that the majority of participants state at least one incorrect rule and that failure is not uncommon. Strikingly, the incorrect rules proposed by participants often express one of the salient features of the initial exemplar (2, 4, 6), such as *even numbers,* or *increasing by two* which it seems difficult for them to eliminate. Based on this observation, there have been many comments in terms of *confirmation bias,* that is, a tendency to look for evidence that examplifies the rule rather than for evidence that contradicts it.

The other task is, on a par with the conditional reasoning task, the most extensively used in the study of reasoning; it is called the *selection task* and is also due to Wason (1966). The participant is presented with four cards and told that each card has a number on one side and a letter on the other side. The four cards displayed show an A, a K, a 4 and a 7, respectively. The participant is then asked to consider the following rule applied to the cards: "If there is an A on one side, then there is a 4 on the other side". The question is: "indicate which of the four cards would need to be turned over in order to decide whether the rule is true or false". The reader who wishes to try to solve the task before going any further should be careful: the task is much more difficult than it seems. With little variation across populations, it is a very robust result that the dominant patterns of selection are (i) the *A* card alone and (ii) the *A* card and *4* card. In terms of individual selections, the *A* card is selected by most people, the *4* card by about one half of the people, and the *7* card by a small minority only. These results have triggered numerous experiments and many comments on human irrationality, as they are at variance with what is considered the normative solution from the falsificationist point of view in inductive logic (and also from the point of view of those who consider the task as a kind of deductive exercice in propositional logic): (i) only the A card and the 7 card are correct choices because they are the only ones that can potentially falsify the rule, and (ii) the 4 card is an incorrect choice because, although it may confirm the rule, it cannot falsify it.

3. An assessment of performance

To the extent that, unlike the case of decision making, the identification of the appropriate normative system to be used can be done without much disagreement, performance is easy to assess. The tasks that have been reviewed were designed to evaluate basic inductive and deductive reasoning ability. At first sight, the overview of performance is contrasted. On the one hand, there is evidence of an overall logical competence demonstrated by the facts that performance is clearly better than chance, and that it is nearly perfect on a number of sub-tasks; but on the other hand, there are response biases, and performance is surprisingly low on many sub-tasks, not to talk of the selection task as a whole. This leaves a sizeable proportion of participants (who, it should be rembered, are mostly University students) apparently failing to apply some basic deductive arguments and inductive principles or methods, a serious blow for proponents of human rationality. But is such a conclusion correct? Before accepting it, it seems necessary to investigate in more detail both the content of the tasks and their demands in terms of the amount of information to be processed.

3.1. The validity of the task: pragmatic analysis.

To begin with, is there not a problem of validity that affects the interpretation of the results on a number of tasks? As we will see, the answer is affirmative. The development of linguistic pragmatics in the seventies in France (Ducrot, 1971, 1972) and in England (Grice, 1975, 1978) made available the concepts necessary for analysis of the communication between experimenter and subject with a view to identifying what the latter actually understands in the experimental situation. This analysis results in determination of the interpretation of the verbal message. In conjunction with the analysis of the representation of the task (what the participants think the experimenter expects them to do) based on task particulars and knowledge of the sollicited population, it is possible to define the input information and what the participant decides to do out of it, and check whether these components coincide with the experimenter's expectations. Only in the affirmative is the task valid. For a general presentation of this approach, see Hilton, 1995; Politzer, 1986, in press; Politzer and Macchi, 2000. Let us now review some of the tasks from this viewpoint.

Firstly, consider conditional reasoning and the two fallacies. Notice that in various contexts (e.g., when uttering promises or permissions) it is often appropriate to interpret a conditional *if P then Q* as *if and only if P then Q* (called *biconditional).* Now under the biconditional interpretation the two arguments concerned are valid. By applying some manipulation, it is possible to avoid the biconditional interpretation (e.g. by adding a premise or by ruling out explicitly the *only if* interpretation): the result is a drop by one half in the rate of "fallacious" responses (Rumain, Connell and Braine, 1983). In brief: (i) about half of the fallacies are cases where it is not the specific argument prepared by the experimenter that is treated by the participant; the latter treats correctly a different argument to which his answer is evaluated as incorrect. (ii) the participant's interpretation of the conditional sentence has nothing to do with formal logic; it has to do with semantics and pragmatics and there is no normative interpretation.

There is even more to say about the fallacies because there is another source of apparent erroneous answer due to a failure by the experimenter to communicate the characteristics of the expected conclusion, namely a deductive conclusion. If the argument is tackled from an abductive point of view (that is, an inference to a plausible justification), the "fallacious" responses are clearly correct. When this is taken into account, most of the remaining errors disappear (Stilgenbauer, 2001).

In summary, the fallacies in conditional reasoning can be explained on the basis of an interpretative phenomenon or a problem of communication; there remains, however, one deductive argument (MT) with a rate of error of about one third to which this explanation does not apply and which can be regarded as a genuine limitation in human inferential abilities.

The immediate inference task also is liable to an analysis from an interpretive point of view. The relevant concept (also applicable to the case of the conditional) is that of *implicature.* An implicature is a proposition that is not part of the literal meaning of a speaker's utterance but is added to it by the hearer on the basis of the context and of mutual knowledge and expectations. So, in ordinary conversation, saying that *some white marbles are small*, often communicates the implicature that *not all white marbles are small*. In standard pragmatic theory, this is based on the assumption that the hearer expects the speaker to be as informative as is appropriate, so that if the speaker knew or believed that all the white marbles are small she would have said so; and consequently, the fact that she has not said so licenses and suggests that she does not know or believe that this is so, hence the interpretation *some, but not all*. As participants in psychological experiments are untrained in logic, they have no reason to construe the sentences literally; rather, it must be accepted that they interpret the sentences presented to them as they do in daily life. It follows that the erroneous responses made to the immediate inferences that have a *some* premise or conclusion demonstrate in fact correct reasoning. Take for instance the inference from *all…* to *some…* which literally requires a 'true' answer. When it is interpreted as an inference from *all…* to *some but not all…* it becomes an inference from a sentence to its contradictory and the answer 'false' becomes the correct one. In brief, for those immediate inferences that have been described, performance can be considered as excellent once the interpretative component has been taken into account.

The pragmatic analysis of both of Wason's tasks is also illuminating. The main source of difficulty is revealed by the consideration of the triple *2,4,6*. It has very salient features; given that it has been specially selected and presented as an instance by the experimenter, participants are thereby invited to assume that its features are relevant. But in fact, these features overdetermine the rule (the numbers need not be even, they need not increase by two, etc., in order to follow the rule actually used which is *three increasing numbers).* In brief, the example is too specific so that it can be considered as pragmatically infelicitous. Recent experimental data suport this analysis (Van der Henst, Rossi and Schroyens, 2002). The whole situation is deceptive and the use of the salient features is no clear evidence of a confirmation bias; the extent to which there is such a bias is object of much debate (Poletiek, 2001).

Similarly, in Wason's selection task, selection of the *4* card and failure to select the *7* card are often considered as evidence that people are unable to look for falsificatory information. However, most research on this task has overlooked a fundamental point: the sentence is presented to ordinary people who interpret it as an ordinary indicative conditional. Developing a linguistic pragmatic approach based on relevance theory (Sperber and Wilson, 1995), Sperber, Cara and Girotto (1995) have argued that participants untrained in formal logic evaluate the truth of the sentence through its testable consequences; this is because the natural way to achieve relevance, that is, to contribute to increase the reasoner's stock of knowledge (at a reasonable cost in terms of mental effort) is to enable him to infer consequences from the statement. There are three ways for a conditional rule to be relevant in this sense. One is to make the highly available inference of Modus Ponendo Ponens: knowing that there is an *A* (select the *A* card) draw the conclusion that there is a *4* (look for the number *4* at the back): this produces the first pattern of response. The second way is to infer that there are joint cases of *A* and *4,* (which is not demanded by formal logic), which suggests to select the *A* card to search for the number *4,* and select the *4* card to search for the letter *A:* this produces the second pattern of response and, importantly, explains the apparent search for confirmation. The last way to render the rule relevant is the most sophisticated: inferring that there are no cases of *A* without a *4;* this produces the correct pattern (select the *A* card to look for 7 and select the *7* card to look for *A)*; it is all the more rare as the *not-4* concept (that is, *7)* is less salient. The authors have provided experimental results that support their approach. In brief, the natural way to interpret the task, far from being the test of a hypothesis, is to chek the elementary consequences of the conditional statement suggested by the use of language. Ironically, the comprehension mechanisms pre-empt any reasoning mechanism, so that the task is not one of propositional logic, or of hypothesis testing, let alone a reasoning task in the strict sense. In a word, the selection task is invalid as a reasoning task and cannot teach us much about human rationality, contrary to the claims made by most commentators. However, it does show that a small minority of people can have a representation of the task that coincides with the expectations of the experimenter.

3.2. The information to process: memory load

Some of the tasks reviewed earlier seem free from interpretive difficulties and yet performance may be mediocre to say the least: in our sample of tasks this is the case of the letter series completion task, of syllogisms, and of MT. We are now dealing with the reasoning process proper that leads to a conclusion. As already noticed, within the same type of task there is a wide variety in difficulty and it is generally agreed that this reflects differences in memory load or in the availability of the appropriate strategies. Leaving aside the latter point which relates to individual characteristics such as learning and expertise, the former point is a task characteristic: different tasks and problems require different levels of working memory capacity and the limitation of this capacity is a fundamental feature of human cognition

Take the completion task. Success requires the identification of a periodicity followed by the discovery of a rule that will enable the generation of the missing letter. In order to understand the main factor of difficulty, the notion of a *list* must be defined. A list is a regular sequence of letters whose members, each in turn, feed in the periods of the series. For instance, in problem 2 the list is the arbitrary sequence *b a* out of which *b* and *a* in turn occupy a place in the periods of three letters. In problem 4, there are two lists, both of which consist of the alphabet. The period is again three letters long, the first list provides two letters *(wx* in the first period, *xy* in the second, etc.) and the second list provides the last letter *(a,* then *b,* etc.) The notion of a list explains the main source of difficulty: when problems are partitioned between easy and difficult, this categorisation correlates very strongly with the number of lists (one versus two). Indeed, the load in working memory differs critically depending on whether the reasoner has to keep track of a place in just one list or in two lists.

Similarly, researchers agree that the load in working memory is the main source of difficulty in deductive reasoning, whatever the theoretical point of view they adopt (that of mental model theory or that of mental rules).

Staying within the limits of the symbolic information processing paradigm, two main, competing theoretical views on deductionexist. Notice that both of them implicitly take some form of standard logic to be normative. The *mental rule* approach (Braine and O'Brien, 1998; Rips, 1994) hypothesizes that the human mind is equiped with a set of core rules (similar to some of the rules that belong to Gentzen's natural deduction system); an inference is performed like a formal proof. One of the main predictions of this view is that the longer the derivation is, the more difficult the inference should be, because the memory load increases with this length. The *mental model* approach hypothesizes that people construct an abstract, internal representation of the meaning of the premises, which may result in one or more models; an inference is performed by considering a putative conclusion (presented to the reasoner, or which (s)he has identified in the model under consideration, and inspecting the other models in an attempt to falsify this putative conclusion: the conclusion follows validly if it holds in all the models. The main prediction is that the more models there are to inspect, the greater the difficulty of the inference, as the memory load increases with this number.

Take for instance Modus Tollendo Tollens *(if P then Q; not Q; therefore not-P).* In mental rule theory, this argument does not belong to the core rules, and therefore it must be derived. Given the first premise *(if P then Q)*, the reasoner supposes *P* and right away infers *Q* by applying MP (which is assumed to belong to the core rules). But *Q* contradicts the second premise *not-Q;* by a rule of *reductio ad absurdum* this allows the negation of the supposed proposition *P,* which yields the conclusion: *not-P.* This lengthy derivation explains why, as reported above, MT is harder than MP. (One should also mention the possible lack of availability to some reasoners of the rather sophisticated strategic rules *supposition* and *reductio ad absurdum*).

In mental model theory, the reasoner must built integrated model(s) of the two premises. The conditional has three models (akin to the three lines of the truth table where it is true): [P   Q] is explicit, but the other two, [¬P   Q] and [¬P ¬Q] are assumed to be implicit (that is, less available and called on as a second resort). MP is immediately solved because the second premise *P* appears in the first model, where *Q* can be found, yielding the conclusion. (The second and third models are somehow irrelevant because they do not accommodate the second premise). To solve MT, some reasoners inspect only the first model, do not recognize ¬*Q* in it, and conclude that nothing follows from the premises. Better reasoners "flesh out" the models, that is, built the other two and notice $\neg Q$ in the last one, which provides an integrated model of the premises, in which $\neg P$ can be found, so yielding the conclusion. The necessity to flesh out the models explains the greater difficulty of MT.

Limitation of working memory capacity is likely to be at the origin of many biases. When an argument is too difficult to handle, the participant uses a fall-back strategy: rather than opting for the commonly available "don't know" option, it may be more appropriate to venture a response based on some superficial feature of the sentences, with the possible bonus that it may coincide with the correct answer. For example, compare the following premise pairs of three syllogisms: (i) all M are P; all S are M. (ii) all P are M; all S are M. (iii) all M are P; all M are S. The conclusion *all S are P* is correctly given by 90% of the participants to the first one and incorrectly by 60% to 80% to the second one (which is invalid) and to the third one (to which the correct response is *some S are P).* If a heuristic such as "conclude with the same quantifier as in the premises" is applied, it is highly successful in the first case but fails in the other two. We now turn to the question of heuristics and biases.

4. Reassessing results in the judgment and decision making domain

The analysis used in section 3.1 based on methodological considerations can be applied to experimental results in other domains. One of these domains covers studies of judgement under uncertainty. The volume edited by Kahneman, Slovic and Tversky twenty years ago provided an impressive catalogue of human shortcomings, such as insensitivity to sample size, misconception of chance, misconception of regression, illusory correlation, overconfidence, attribution error, misunderstnding of conditional probability, base rate neglect, conjunction fallacy, and many more besides. Now, many of these limitations simply reflect that lay people do not possess the sophisticated knowledge of statistician and mathematician; this is similar to ordinary people's failing tests in elementary dynamics that require the correct concepts of mass, velocity and acceleration, resorting instead to their intuition based on a world with friction. Some other results concern more elementary skills or principles, and it can be shown that failure is not so general as claimed by the first investigators.

Consider the two problems that have been the most studied, namely the *conjunction fallacy* and the *base rate neglect.* The first one refers to a classic paradigm (the *Linda problem*, Tversky and Kahneman, 1983) in which participants are presented with a character through a short scenario, and then asked to estimate the probability of two events involving the character, B (typical of the character) and A (not typical). Over eighty percent of the participants estimate p(A&B) to be greater than p(A), a result the authors explained on the basis of the representativeness heuristic (B is more representative of the character than A). However, the pragmatic analysis of the task shows that participants are implicitly invited to compare p(A&B) with p(A&not-B). It can be estimated that the fallacy about one quarter of the participants genuinely commit the fallacy (Politzer and Noveck, 1991; Dulany and Hilton, 1991). Similarly, the neglect of the base rate information seems to be much less general than initially claimed. Koehler (1996) reviews two decades of research and concludes that on the contrary base rates are almost always used. In the classic paradigm (*the lawyer-engineer problem,* Kahneman and Tversky, 1973) participants are told about a sample of two kinds of professionals, the engineer to lawyer ratio being 70/30 in one condition and 30/70 in the other. They are presented with a personality description typical of an engineer, supposed to have been drawn at random, and then asked to estimate the probability that the person described is an engineer. Because the estimates in the two conditions differed by only 5%, it was concluded that base rates (assumed to be given by the ratios) were neglected, even though this difference was statistically significant. In fact, various replications have consistently produced similar results with a somewhat greater effect, which is the important point, showing that on the average there *is* sensitivity to base rates. Furthermore, as argued by Schwarz, Strack, Hilton, and Naderer (1991), the exploitation of the information by participants depends on the inferred communicative intention of the experimenter; that is, emphasis put on the description invites participants to consider the task as a request to exhibit psychological intuition rather than mathematical skill, and therefore to give more weight to the related information.

5. Two kinds of rationality?

The upshot of the foregoing analysis is that in reasoning experiments a majority of people perform fairly highly provided (i) the task is not too demanding in terms of memory load, and (ii) they have the correct representation of the task. The former point relates to a notion introduced nearly half a century ago: Simon (1957) initially introduced the concept of *bounded rationality* in reference to humans' limited capacity for processing information. This leaves us with the question, What factors lead an individual towards a correct or an incorrect task representation?

In recent years, evidence has been obtained from various sources which suggests the existence of two different thinking processes. Evans and Over (1996) in particular distinguish two kinds of rationality:

> *Rationality$_1$:* Thinking, speaking, reasoning, making a decision, or acting in a way that is generally reliable and efficient for achieving one's goals.
> *Rationality$_2$:* Thinking, speaking, reasoning, making a decision, or acting when one has a reason for what one does sanctioned by a normative theory.

A number of authors have made a similar distinction; they share the idea of two systems of reasoning that can be characterised as follows: one is automatic, rapid, associative, contextualised, and does not require much mental load in terms of memory and computation; the other is controlled, slow, rule-based, decontextualised, individualised, and is much more demanding (see Stanovich and West, 2000 for a detailed comparison). Whereas the second system is more linked with formal education, the first system is the product of experience (some view it as an adaptation in the evolutionary sense): it operates successfully and efficiently in real world situations where its use is basically rational in the first sense. This conceptualisation concerns not only the reasoning domain considered earlier but also judgment and decision making. It is argued that the various heuristics described in this domain are instances of the application of the first system to laboratory tasks, and the associated biases are the consequence of applying them as if they were real-life situations. Stanovitch and West have observed a correlation between success at the laboratory tasks and achievement on tests of intelligence, which they interpret as indicating that those who score higher in analytic intelligence, which is linked to the second system, are more prone to apply this system in the laboratory, hence a representation of the task that fits with that of the experimenter. An alternative, not incompatible view, is that those who have been more exposed to formal teaching have developped metacognitive skills and knowledge which are crucial to determine the correct representation of the task; they have also learned methods and strategies incidentally if not purposively and consequently have better chance of succeeding at the laboratory tasks.

In conclusion, participants in laboratory experiments on reasoning (and on judgment and decision making too) usually tackle the task as if the communication was similar to an ordinary conversation in everyday life, to which they react using whatever logical skills they possess; if the sentences are liable to an interpretation which differs from the literal meaning, this may lead them to a formally incorrect response (e.g. responding by 'false' to an inference from *all…* to *some…*) or to a correct response (e.g. responding by 'false' to an inference from *all…* to *no…);* they are rational both in the first and the second senses. However, the educational background of some participants gives them a hint that they should approach the task as a piece of scientific communication, in which every word of the instructions has to be carefully considered, the appropriate interpretation of the sentences and terms of the problems is given a literal meaning, and the information provided is necessary and sufficient, in brief a game where, from the outset, they exhibit rationality in the second sense. Finally, the task may exceed the information processing capacity of the participants, whatever their approach to it, or they may not possess the right method or strategy, in which case they often resort to a heuristic, resulting in biases and erroneous answers (e.g. responding by 'true' to an inference from *some A are not B* to *some B are not A* based on the heuristic 'assume symmetry'); using a heuristic as a last resort is rational in the first sense.

6. Conclusion

Experimental work has demonstrated the existence of normatively incorrect behaviour in both the domain of reasoning and that of judgment and decision making. For practical purposes, the significance of this finding depends crucially on the extent to which such behaviour occurs in real-life situations. For one to reach a conclusion, several conditions should be satisfied. One is that there should be agreement on the appropriateness of the norm being used (see Baratgin, in press, for a discussion of probability judgement from this point of view). Another condition is that the experimental results should have ecological validity. Both requirements have been the object of much debate, especially the former in the domain of probabilistic reasoning and decision making, and a definite answer seems premature. The present paper has focused on yet a third condition, namely the methodological soundness of the experiments. It has been shown that, after a pragmatic re-examination of the experiments, the performance that is revealed is often much better than it first appeared to be. Therefore, in all likelihood, the results of some famous experiments showing an overwhelming proportion of a population violate elementary laws of logic or probability, and consequently basic norms of rationality, are very much exaggerated. Now, those who wish to develop theory and practice on the basis of an assumed human rationality need an answer to the question, What is the "true" extent of the defective behaviour? That a few percent of a population are concerned may be relatively harmless; that one third is may be devastating. Unluckily, this chapter must end on a very familiar note: much more research is needed to answer the question with precision.

Answers to the syllogisms:
(i) no smoker is a player
(ii) some players are not smokers

References

Baratgin, J. (in press). Is the human mind definitely not Bayesian? A review of the various arguments. *Current Psychology of Cognition/Cahiers de Psychologie Cognitive.*

Braine, M. D. S. , & O'Brien, D. P. (1998). *Mental logic.* Mahwah, N.J.: Lawrence Erlbaum.

Ducrot, O. (1972). *Dire et ne pas dire.* Paris: Hermann.

Ducrot, O. (1973). *La preuve et le dire.* Tours: Mame.

Dulany, D.E., & Hilton, D.J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition, 9,* 85-110.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences.* London: Erlbaum.

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

Evans, J.St.B.T., & Over, D.E. (1996). *Rationality and reasoning.* Hove: Psychology Press.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and Semantics.* Vol 3: Speech acts. N. Y.: Academic Press.

Grice, H.P. (1978). Further notes on logic and conversation. In P. Cole (Ed.), *Syntax and Semantics.* Vol 9: Pragmatics N.Y.: Academic Press.

Hilton, D. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118,* 248-271.

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

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction.* Hove & London: L. Erlbaum.

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

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237-251.

Kotovsky, K., & Simon, H.A. (1973). Empirical tests of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology, 4,* 399-424.

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

Poletiek, F.H. (2001). *Hypothesis-testing behaviour.* Hove: Psychology Press.

Politzer, G. (1986). Laws of language use and formal logic. *Journal of Psycholinguistic Research, 15,* 47-92.

Politzer, G. (1990). Immediate deduction between quantified sentences. In K.J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds.), *Lines of thinking. Reflections on the psychology of thought,* (pp.85-97). London: Wiley.

Politzer, G. (in press). Reasoning, judgment and pragmatics. In I. Noveck & D. Sperber (Eds.), *Experimental Pragmatics.* London: Palgrave.

Politzer, G., & Macchi, L. (2000). Reasoning and pragmatics. *Mind and Society, 1,* 73-93.

Politzer, G., & Noveck, I.A. (1991). Are conjunction rule violations the result of conversational rule violations? *Journal of Psycholinguistic Research, 20,* 83-103.

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

Rumain, B., Connell, J., & Braine, M.D.S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. *Developmental Psychology, 19,* 471-481.

Schwarz, N., Strack, F., Hilton, D., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: the contextual relevance of "irrelevant" information. *Social Cognition, 9,* 67-84.

Simon, H.A. (1957). *Models of man.* New York: Wiley.

Simon, H.A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review, 70,* 534-546.

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

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition.* 2nd edition. London: Blackwell.

Stanovich, K.E., & West, R.F. (2000). Individual differences in reasoning: implications for the rationality debate? *Behavioral and Brain Sciences, 23,* 645-726.

Stilgenbauer, J.-L. (2001). *Erreur de raisonnement et syllogismes conditionnels: une approche abductive.* [Reasoning error and conditional syllogisms: an abductive approach]. D.E.A. dissertation. Ecole des Hautes Etudes en Sciences Sociales, Paris.

Van der Henst, J.-B. , Rossi, S. , & Schroyens, W. (2002). When participants are not misled they are not so bad after all : A pragmatic analysis of a rule discovery task. *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Mahwah, N.J.: Lawrence Erlbaum.

Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129-140.

Wason, P.C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in Psychology.* Vol. 1. Pelican Book.