ESSAYS IN COGNITIVE PSYCHOLOGY

RATIONALITY AND REASONING

Jonathan St.B. T. Evans
and David E. Over

RATIONALITY AND REASONING

This page intentionally left blank

Rationality and Reasoning

Jonathan St.B.T. Evans
University of Plymouth, UK

David E. Over
University of Sunderland, UK

Psychology Press
Taylor & Francis Group
HOVE AND NEW YORK

Transferred to Digital Printing 2011

Psychology Press, Publishers
27 Church Road
Hove
East Sussex, BN3 2FA
UK

Psychology Press is an imprint of the Taylor & Francis Group; an Informa business

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 978-0-86377-438-6 (Hbk)

Cover by Clearwater Design, Hove.

Publisher's Note
The publisher has gone to great lengths to ensure the quality of this reprint but points out that some imperfections in the original may be apparent.

Contents

Foreword

There has been a dramatic increase, on both sides of the Atlantic, in contributions to the psychology of reasoning: research into how people actually think rather than how they ought to think. It has even justified a new journal, *Thinking & Reasoning*, devoted exclusively to this topic. Incidentally, the editor of the journal is the first author of this book. It is of passing interest to note that the origins of this research have been due to luck and play, not, as some authorities assume, a concerted attack on the grasp of deductive reasoning. What started as a perplexing puzzle has now been polished and consolidated into a wide variety of complex theories.

The bored psychologist glancing at the shelves marked "cognition" in the local bookshop, might catch sight of this book and inwardly exclaim, "Not another book on reasoning!". Such a dismissive reaction would be plausible, but entirely wrong. In my view, there are at least three features in this book which merit more than a cursory glance.

First, the book is a collaborative effort between two people who have different backgrounds in psychology and philosophy. Jonathan Evans is a distinguished experimental psychologist whose work has spearheaded research on reasoning for some time. David Over was originally a philosopher who has increasingly worked with psychologists on decision theoretic analyses of what had been seen as pure problems in reasoning. His papers with Ken Manktelow have helped to illuminate deontic versions of the Wason selection task. Even more recently, he and Jonathan Evans have started to integrate work on reasoning and decision making in new ideas about indicative versions of the task, the most formidable and intractable of all the problems in the field.

What has been achieved by this collaboration is a liberation from the conventional wisdom of the dominant group, and an appeal to regard reasoning behaviour in a wider focus. Johnson-Laird and I have always

invoked propositional logic as the normative standard for judging performance, but the present authors distinguish between "Rationality$_1$" which is basically a "generally reliable" way of achieving goals, and "Rationality$_2$" which is achieving a goal sanctioned by a normative theory (e.g. logic).

The second reason for commending this book is that it draws not merely on the "reasoning" literature but on the "decision making" literature. Decision making is concerned intrinsically with probability rather than deduction. It has acquired its own sophisticated mathematical methodology, but to invoke both traditions to tackle the same empirical data may be regarded as a daring step. We like to keep our nests immaculate, and not muddied by a different tradition, even if we are totally unaware of this penchant.

A third point of interest lies in a study in which I was privileged to work with Evans (Wason & Evans, 1975), discussed in the final chapter of this book. Its results show a marked discrepancy between the causes of people's actions or decisions and the explanations that they offer for them. This has similarities to the annotations of a chess master which claim to be objective but are frequently rationalisation of a line of play already accepted (or rejected). The concept of rationalisation originates, of course, in a psychoanalytic context where it is regarded as justifying unconscious wishes, demands, etc. In such cases the choice, or decision, is invested with prior meaning or emotion. However, the results which Evans and I obtained in two experiments were based on that trivial material (circles, triangles, etc.) so beloved by the cognitive psychologist. Furthermore, the written protocols in our experiment suggest that what looked like a fluctuation of insight from one test to another, was really governed by a prior non-verbal choice. While the limits of such "dual processing" have not yet been fully established, the current volume contains new and valuable discussion of the role of tacit and explicit thinking.

Not all of us in the firm will agree with everything claimed by the authors. How could it be otherwise? For instance, I am inclined to agree with some "criticisms" recently made about an experiment of mine published nearly 40 years ago, but such criticisms overlook an unanticipated, qualitative result which makes the study a unique example of obsessional thinking under controlled conditions. However, this was a "one-off" experiment which has led only to methodological debate. Nevertheless, I hope enough has been said to show that this book is a fertile ground for others interested in cognition and epistemology. May it meet with the success it deserves.

Peter Cathcart Wason
Oxford, February 1996

Preface and acknowledgements

Readers may be interested to know how a collaborative book between a psychologist and a philosopher living some 400 miles apart came to be written. The first author (JE) has been active in researching the psychology of reasoning since the early 1970s and has made a particular study of biases in reasoning, the subject of an earlier book in this series (Evans, 1989). To his continuing puzzlement he found that many authors in the field were attributing to him a claim that such biases implied that human beings were irrational. Not only were such claims absent from JE's publications, but some pretty clear statements to the contrary had actually been provided. For example:

> The view that I wish to argue here is that errors of thinking occur because of, rather than in spite of, the nature of our intelligence. In other words, they are an inevitable consequence of the way in which we think and a price to be paid for the extraordinary effectiveness with which we routinely deal with the massive information-processing requirements of everyday life. (Evans, 1989, p.111.)

The second author (DO) is a philosopher who has been active in the psychology of reasoning for some years and who regularly defies the rules of the philosophers' union by running experiments. His interest developed through collaboration with the psychologist Ken Manktelow,

and, in 1993, he and Ken edited a volume on rationality to which a number of leading researchers, including JE, were asked to contribute (Manktelow & Over, 1993). In his chapter, JE took the opportunity both to respond to some reviewers of the 1989 book who had criticised it for avoiding the issue of rationality, and to deal with the broader question of what implications research on reasoning biases has for human rationality. This led to the first presentation of the distinction between two definitions of rationality—rationality$_1$ and rationality$_2$—which are discussed at length in this book. The basic idea is that one may be rational in terms of achieving personal goals (rationality$_1$) without being rational in the sense of conforming to a normative system such as logic (rationality$_2$). DO had been working for some time on the idea that behaviour on some reasoning tasks could be better viewed in terms of decision processes rather than as logical reasoning. JE's distinction thus struck an immediately sympathetic chord and instigated a series of discussions that led first to some collaborative papers and eventually to the current volume.

With a title like *Rationality and Reasoning*, readers might wonder whether this book is intended as a work in philosophy or psychology. The answer is both. The primary focus of the work is on the interpretation of the psychological literature on reasoning and decision making—hence publication in this "Essays in Cognitive Psychology" series. However, we deal also with philosophical issues and bring these to bear on the interpretation of the psychological evidence. In the early part of the book we concentrate mostly on conceptual issues, attempting to clarify thinking about the concept of rationality in psychological research and exploring the limitations of normative theoretical analysis. As the book progresses, we increasingly develop our psychological theory of reasoning and decision processes. Although our differing backgrounds were obviously helpful in this endeavour, we would like to stress that every part of this book has been drafted or redrafted by each of us. We both wrote all of it.

Although a relatively short book, the curent volume was written over the deliberately protracted period of two years in order to allow full consideration of some very complex questions. During this period both authors benefited from periods of fellowship leave, which made the task a lot easier than it might have been. We would like therefore to acknowledge our employers—the Universities of Plymouth and Sunderland—for their support in this enterprise. DO would also like to acknowledge the Psychology Department of Princeton University where he was a Visiting Fellow while on leave and where he benefited particularly from discussions with Phil Johnson-Laird, Danny Kahneman, and Dick Jeffrey. JE spent part of his leave as Visiting

Professor to CREPCO in the University of Provence, France where he had many relevant discussions with Paolo Legrenzi and Vittorio Girotto.

There are a number of other colleagues whose assistance we would like to acknowledge. First, we must especially thank those who read and commented on the entire book in draft manuscript form: Dan Sperber, Paolo Legrenzi, Jonathan Lowe, and John Clibbens. We made significant changes in response to each of these commentators. Next, we acknowledge that our thinking about these issues has benefited from discussions with various collaborators in our recent work on the psychology of reasoning and decision making—in particular Ken Manktelow, Rosemary Stevenson, Steve Newstead, Ruth Byrne, Simon Handley, and David Green. Finally, there are those with whom we have not formally collaborated, but with whom we have discussed some of the ideas that feature in this book; in particular we would like to mention Kris Kirby, Patricia Cheng, Mike Oaksford, and Nick Chater.

This page intentionally left blank

CHAPTER ONE

Rationality in reasoning

Reasoning and decision making are topics of central importance in the study of human intelligence. Reasoning is the process by which we can apply our vast stores of knowledge to the problem at hand, deducing specific consequences from our general beliefs. Reasoning also takes place when we infer the general from the specific, by formulating and then testing new ideas and hypotheses. Rules for correct reasoning have been laid down by great thinkers in normative systems (principally logic and probability theory) and it is tempting to define, and evaluate, human rationality by referring only to these rules. However, we shall argue in this book that this approach is mistaken. The starting point for any understanding of human rationality should be behavioural: we must ask how decisions taken and actions performed serve the goals of the individual. Formulating and making use of logical and other rules has always had to rest on a more fundamental human ability to achieve behavioural goals.

The psychology of thinking has a long history, but the past 25 years or so have witnessed an explosion of research effort in the areas of decision making, judgement and reasoning, with many hundreds of experiments reported in the psychological literature. These studies are reviewed in detail in several recent textbooks (e.g. Baron, 1988; Evans, Newstead, & Byrne, 1993a; Garnham & Oakhill, 1994) and we will make no attempt to repeat the exercise here. In this book our purposes are theoretical and integrative: we seek to make sense of research in these

vast literatures and to resolve some central theoretical issues of importance concerning rationality and the nature of human thinking. Hence our discussion of the published studies will be highly selective, although we believe focused on findings that are both important and representative of the area as a whole.

The aims of this book are three-fold. First, we address and attempt to resolve an apparent paradox of rationality that pervades in these fields. The issue of rationality is central to the first two chapters and underpins much of our later discussion. Next, we seek to achieve integration between the study of reasoning and decision making. Despite some recent efforts to bridge the gap, research in these two areas has proceeded largely in isolation. It seems to us that the mental processes of reasoning and decision making are essentially similar, although we shall see how an emphasis on rule-following as the basis of rationality has rendered this resemblance less than self-evident. Finally, we shall present a dual process theory of thinking which advances understanding of the phenomena we discuss and the psychological mechanisms underlying the kinds of rationality that people display in their reasoning and decision making.

THE RATIONALITY PARADOX

The human species is far and away the most intelligent on earth. Human beings are unique in their cognitive faculties—for example, their possession of an enormously powerful linguistic system for representing and communicating information. They have learned not only to adapt to the environment but to adapt the environment to suit themselves; and they have organised vastly complex economic, industrial, and political systems. They have also developed a capacity for abstract thinking that has enabled them, among other things, to create logic and other normative systems for how they ought to reason.

What happens when representatives of this highly intelligent species are taken into the psychological laboratory to have their processes of thinking, reasoning, and decision making studied by psychologists? The surprising answer is that people seem to make mistakes of all kinds, as judged by the normative rules that human beings have themselves laid down. Many of these rule violations are systematic and arise in cases where a bias is said to exist. Some psychologists use this term as a necessarily pejorative one, but for us it will be descriptive, meaning only a departure from an apparently appropriate normative system. We are not talking here about minor aspects of human performance: systematic deviations from normative principles have been identified and reported

in many hundreds of published studies within the past twenty years alone. Although we discuss the issues in general terms in this chapter, a number of specific examples of reasoning and decision biases will be discussed throughout this book. Lest this chapter be too abstract, however, we present a single example of the kind of thing we are talking about.

In syllogistic reasoning tasks, subjects are presented with two premises and a conclusion. They are instructed to say whether the conclusion follows logically from the premises. They are told that a valid conclusion is one that must be true if the premises are true and that nothing other than the premises is relevant to this judgement. Suppose they are given the following problem:

1.1 No addictive things are inexpensive
Some cigarettes are inexpensive
Therefore, some addictive things are not cigarettes

On the basis of the information given, this syllogism is invalid. In other words, the conclusion does not necessarily follow from the premises. The class of cigarettes might include all addictive things, thus contradicting the conclusion. Of course, those cigarettes that were inexpensive would not be the ones that were addictive, but this is quite consistent with the premises. However, the majority of subjects given problems like 1.1 state erroneously that the conclusion does follow logically from the premises (71% in the study of Evans, Barston, & Pollard, 1983). Now suppose the problem is stated as follows:

1.2 No cigarettes are inexpensive
Some addictive things are inexpensive
Therefore, some cigarettes are not addictive things

The logical structure of 1.2 is the same as 1.1; all we have done is to interchange two of the terms. However, with problems of type 1.2 very few subjects say that the conclusion follows (only 10% in the study of Evans et al., 1983). What is the critical difference between the two? In the case of 1.1 the conclusion is believable and in the case of 1.2 it is not believable. This very powerful effect is known as "belief bias" and is discussed in detail in Chapter 5. It is clearly a bias, from the viewpoint of logic, because a feature of the task that is irrelevant given the instructions has a massive influence on judgements about two logically equivalent problems.

There seems to be a paradox. On the basis of their successful behaviour, human beings are evidently highly intelligent. The

psychological study of deduction, on the other hand, appears to suggest that they are illogical. Although some authors in research on biases have been careful to qualify their claims about human behaviour, others have made fairly strong claims that their work shows people to be irrational (see Lopes, 1991, which discusses a number of examples). It is perhaps not surprising that research on biases in reasoning and judgement has come under close scrutiny from philosophers and psychologists who simply cannot accept these findings at face value and who take exception to the inferences of irrationality that are often drawn from the studies concerned. These criticisms mostly come from authors who take human rationality to be obvious, for the reasons outlined above, and who therefore conclude that there is something wrong with the research or its interpretation.

Evans (1993a) has discussed the nature of this criticism of bias research in some depth and classified the arguments into the following three broad groupings:

- the normative system problem;
- the interpretation problem; and
- the external validity problem.

The first major critique of bias research by a philosopher was that of Cohen (1981) whose paper includes examples of all three types of argument. The normative system problem, as Cohen discusses it, is that the subject may be reasoning by some system other than that applied in judgement by the experimenter. For example, psychologists studying deductive reasoning tend to assume a standard logic—such as extensional propositional logic—as their normative framework, whereas many other logics are discussed by logicians in the philosophical literature. Cohen (1982) suggested further that people might be using an old Baconian system of probability based on different principles from modern probability theory. We find this suggestion implausible—substituting, as it does, one normative system for another. However, the idea that rationality is personal and relative to the individual is important in our own framework as we shall see shortly.

A different slant on the normative system problem is the argument that conventional normative theories cannot be used to assess rationality because they impose impossible processing demands upon the subject. We would not, for example, describe someone as irrational because they were unable to read the text of a book placed beyond their limit of visual acuity, or because they could not recall one of several hundred customer addresses, or were unable to compute the square root of a large number in their heads. For this reason, Baron (1985)

distinguishes normative theories from prescriptive theories. The latter, unlike the former, prescribe heuristics and strategies for reasoning that could be applied by people within their cognitive processing capabilities. For example, people cannot be expected to internalise probability theory as an axiomatic system and to derive its theorems, but they can learn in general to take account of the way in which the size and variability of samples affects their evidential value.

In the case of deductive reasoning, this type of argument has been proposed in several recent papers by Oaksford and Chater (e.g. 1993, 1995). They point out that problems with more than a trivial number of premises are computationally intractable by methods based on formal logic. For example, is is known that to establish the logical consistency of n statements in propositional logic requires a search that increases exponentially with n. Oaksford and Chater go on to argue that the major theories of deductive reasoning based on mental logic and mental models (discussed later) therefore face problems of computational intractability when applied to non-trivial problems of the sort encountered in real life, where many premises based on prior beliefs and knowledge are relevant to the reasoning we do. In this respect the argument of Oaksford and Chater bears also upon the external validity problem, also discussed later.

The interpretation problem refers to the interpretation of the problem by the subject, rather than the interpretation of the behaviour by the psychologist. The latter is a problem too, but one which belongs under the third heading, discussed below. The interpretation argument has featured prominently in some criticism of experimental research on deductive reasoning. For example, in a very influential paper, Henle (1962) asserted that people reason in accordance with formal logic, despite all the experimental evidence to the contrary. Her argument is that people's conclusions follow logically from their personalised representation of the problem information. When the conclusion is wrong, it is because the subject is not reasoning from the information given: they may, for example, ignore a premise or redefine it to mean something else. They might also add premises, retrieved from memory. Henle illustrates her argument by selective discussion of verbal protocols associated with syllogistic reasoning. There are some cases, however, in which her subjects appear to evaluate the conclusion directly without any process of reasoning. These she classifies as instances of "failure to accept the logical task".

Another version of the interpretation problem that has received less attention than it deserves is the argument of Smedslund of a "circular relation between logic and understanding" (see Smedslund, 1970, for the original argument and 1990 for a recent application of it).

Smedslund argues that we can only decide if someone is reasoning logically if we presume that they have represented the premises as intended. Conversely, we can only judge their understanding of the problem information if we assume that they have reasoned logically. Smedslund's surprising conclusion from his discussion of this circularity is that "the only possible coherent strategy is always to presuppose logicality and regard understanding as a variable". This argument was scrutinised in detail by Evans (1993a) who refuted it by discussion of the specific example of conditional inference. He showed that subjects' reasoning in such cases is not logically consistent with any interpretation that can be placed upon the conditional sentence and nor is there logical consistency between reasoning on one problem and another.

Perhaps the most potentially damaging critique of bias research is that based on the external validity problem. In its least sympathetic form, as in Cohen's (1981) paper, the argument can aim to undermine the value of the research fields concerned on the basis that they study artificial and unrepresentative laboratory problems. Consider, for example, the Wason selection task (Wason, 1966), which we discuss in some detail later in this book. Devised as a test of hypothesis testing and understanding of conditional logic, this problem is solved—according to its conventional normative analysis—by less than 10% of intelligent adult subjects, and has become the single most studied problem in the entire reasoning literature (see Evans et al., 1993a, Chapter 4, for a detailed review). Cohen attempted to dismiss the phenomenon as a "cognitive illusion", analogous to the Muller-Lyer illusion of visual perception. If he is right, then many researchers have chosen to spend their time studying a problem that is wholly unrepresentative of normal thinking and reasoning and that presents an untypically illogical impression of human thought. We disagree with Cohen, but we will nevertheless consider in some detail how performance on this particular task should be interpreted. Where we will agree with him is in rejecting the notion that the selection task provides evidence of irrationality. However, unlike Cohen we believe that study of this task has provided much valuable evidence about the nature of human thought.

Other aggressive forms of the external validity argument include suggestions that bias accounts are proposed to accord with fashion and advance the careers of the psychologists concerned and that researchers create an unbalanced picture by citing disproportionately the results of studies that report poor reasoning (see Berkeley & Humphreys, 1982; Christensen-Szalanski & Beach, 1984). A milder version of the argument has been presented by such authors as Funder (1987) and

Lopes (1991) who, like us, are sympathetic to the research fields but concerned by interpretations of general irrationality that are placed upon them. Experiments that are designed to induce errors in subjects' performance are valuable in advancing our theoretical understanding of thought processes. It is a mistake, however, to draw general inferences of irrationality from these experimental errors. As an analogy, consider that much memory research involves overloading the system to the point where errors of recall will occur. This provides useful experimental data so that we can see, for example, that some kinds of material are easier to recall than others, with consequent implications for the underlying process. Such research is not, however, generally used to imply that people have bad and inadequate systems of memory. So why should explorations of cognitive constraints in reasoning be taken as evidence of poor intelligence and irrationality?

Our own theoretical arguments stem from an attempt to resolve the problems outlined in this section and to address some of the specific issues identified. In doing this we rely heavily upon our interpretation of a distinction between two forms of rationality, first presented by Evans (1993a) and by Evans, Over, and Manktelow (1993b).

TWO KINDS OF RATIONALITY

Human rationality can be assessed in two different ways: one could be called the personal and the other the impersonal. The personal approach asks what our individual goals are, and whether we are reasoning or acting in a way that is usually reliable for achieving these. The impersonal approach, in contrast, asks whether we are following the principles of logic and other normative theories in our reasoning or decision making. Flanagan (1984, p. 206) gives a good statement of an impersonal view of rationality:

> Often rationality is taken as equivalent to logicality. That is, you are rational just in case you systematically instantiate the rules and principles of inductive logic, statistics, and probability theory on the one hand, and deductive logic and all the mathematical sciences, on the other.

Taking this approach, we would say that people are rational if they have reasons for what they believe or do that are good ones according to logic or some other impersonal normative system. But this way of looking at rationality should be combined with a personal view, which sees an individual's mental states or processes as rational if they tend