



hypothetical thinking

dual processes in
reasoning and judgement

JONATHAN St. B. T. EVANS

Essays in Cognitive Psychology

HYPOTHETICAL THINKING

Hypothetical thought involves the imagination of possibilities and the exploration of their consequences by a process of mental simulation. Using a recently developed theoretical framework called Hypothetical Thinking Theory, Jonathan St B. T. Evans provides an integrated theoretical account of a wide range of psychological studies on hypothesis testing, reasoning, judgement and decision making.

Hypothetical thinking theory is built on three key principles, implemented in a revised and updated version of Evans' well-known heuristic-analytic theory of reasoning. The central claim of this book is that this theory can provide an integrated account of some apparently very diverse phenomena including confirmation bias in hypothesis testing, acceptance of fallacies in deductive reasoning, belief biases in reasoning and judgement, biases of statistical judgement and a number of characteristic findings in the study of decision making. The author also provides broad ranging discussion of cognitive biases, human rationality and dual-process theories of higher cognition.

Hypothetical Thinking draws on and develops arguments first proposed in Evans' earlier work from this series, *Bias in Human Reasoning*. In the new theory, however, cognitive biases are attributed equally to analytic and heuristic processing and a much wider range of phenomena are reviewed and discussed. It will therefore be of great interest to researchers and post-graduates in psychology and the cognitive sciences, as well as to undergraduate students looking for a comprehensive review of current work on reasoning and decision making.

ESSAYS IN COGNITIVE PSYCHOLOGY

North American Editors:

Henry L. Roediger, III, *Washington University in St. Louis*

James R. Pomerantz, *Rice University*

European Editors:

Alan D. Baddeley, *University of York*

Vicki Bruce, *University of Edinburgh*

Jonathan Grainger, *Université de Provence*

Essays in Cognitive Psychology is designed to meet the need for rapid publication of brief volumes in cognitive psychology. Primary topics include perception, movement and action, attention, memory, mental representation, language and problem solving. Furthermore, the series seeks to define cognitive psychology in its broadest sense, encompassing all topics either informed by, or informing, the study of mental processes. As such, it covers a wide range of subjects including computational approaches to cognition, cognitive neuroscience, social cognition, and cognitive development, as well as areas more traditionally defined as cognitive psychology. Each volume in the series makes a conceptual contribution to the topic by reviewing and synthesizing the existing research literature, by advancing theory in the area, or by some combination of these missions. The principal aim is that authors provide an overview of their own highly successful research programme in an area. Volumes also include an assessment of current knowledge and identification of possible future trends in research. Each book is a self-contained unit supplying the advanced reader with a well-structured review of the work described and evaluated.

Titles in preparation

Gernsbacher: *Suppression and Enhancement in Language Comprehension*

Park: *Cognition and Aging*

Mulligan: *Implicit Memory*

Surprenant & Neath: *Principles of Memory*

Brown: *Tip-of-the-tongue Phenomenon*

Recently published

Gallo: *Associative Illusions of Memory*

Sarris: *Rational Psychophysics in Humans and Animals*

Cowan: *Working Memory Capacity*

McNamara: *Semantic Priming*

Brown: *The Déjà Vu Experience*

Coventry & Garrod: *Seeing, Saying, and Acting*

Robertson: *Space, Objects, Minds, and Brains*

Cornoldi & Vecchi: *Visuo-spatial Working Memory and Individual Differences*

Sternberg *et al.*: *The Creativity Conundrum*

Poletiek: *Hypothesis-testing Behaviour*

Garnham: *Mental Models and the Interpretations of Anaphora*

For continually updated information about published and forthcoming titles in the Essays in Cognitive Psychology series, please visit: www.psypress.com/essays

Hypothetical Thinking

Dual Processes in Reasoning and Judgement

Jonathan St B. T. Evans

First published 2007 by Psychology Press
27 Church Road, Hove, East Sussex BN3 2FA

Simultaneously published in the USA and Canada
by Psychology Press
270 Madison Avenue, New York, NY 10016

This edition published in the Taylor & Francis e-Library, 2007.

“To purchase your own copy of this or any of Taylor & Francis or Routledge’s collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk.”

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

© 2007 Psychology Press

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

This publication has been produced with paper manufactured to strict environmental standards and with pulp derived from sustainable forests.

British Library Cataloguing in Publication Data

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

Library of Congress Cataloging in Publication Data

Evans, Jonathan St. B. T., 1948–

Hypothetical thinking: dual processes in reasoning and judgement/

Jonathan St. B. T. Evans. – 1st ed.

p. cm. – (Essays in cognitive psychology)

Includes bibliographical references and index.

ISBN-13: 978-1-84169-660-7 (hard back)

ISBN-10: 1-84169-660-9 (hard back)

1. Reasoning (Psychology) 2. Judgment. 3. Thought and thinking. I. Title.
BF442.E93 2007

153.4 – dc22 2006037166

ISBN 0-203-94748-7 Master e-book ISBN

ISBN: 978-1-84169-660-7 (Print Edition)

Contents

Foreword and acknowledgements	vii
1. Introduction and theoretical framework	1
2. Hypothesis testing	25
3. Suppositional reasoning: if and or	49
4. The role of knowledge and belief in reasoning	79
5. Dual processes in judgement and decision making	107
6. Thinking about chance and probability	131
7. Broader issues	155
References	177
Author index	197
Subject index	203

Foreword and acknowledgements

This book is my third contribution to the *Essays in Cognitive Psychology* series, following *Bias in Human Reasoning* (1989) and *Rationality and Reasoning* (with David Over, 1996a). It has some features in common with these earlier books, including an attempt to integrate work from the psychology of reasoning with that on judgement and decision making, and a framing within dual-process theory. However, the current volume represents, I hope, considerably more than an update of these previous works. Of the three books, it is the most theoretically ambitious. I present here a recently developed theory of hypothetical thinking, including a revised and extended version of the heuristic–analytic theory of reasoning. The claim on which the book is based is that phenomena on a wide range of apparently diverse cognitive tasks in the psychological literatures on hypothesis testing, reasoning, statistical judgement and decision making can be understood with reference to a common and relatively simple set of principles. In support of this, I present an extended review and discussion of the relevant studies.

As befits the series, this book is an extended essay and not a textbook. For this reason, I have given more weight at times to discussion of studies that I feel are particularly relevant to the theoretical objectives of the book, including those run in my own laboratory. However, the book includes a fairly comprehensive review of the main findings in the fields covered and should hence prove useful also as a broad introduction to these topics.

I am indebted in this work to my two closest collaborators of recent years, David Over and Simon Handley. In particular, David and Simon helped

me to develop the three principles of hypothetical thinking that form the foundation of hypothetical thinking theory. They have also collaborated with me on a number of experimental investigations of the key phenomena in the study of hypothetical thinking. In addition, I am grateful to several colleagues who read and criticized a draft manuscript of this book, including Keith Stanovich, Valerie Thompson, Shira Elqayam and an anonymous reviewer. The book was certainly improved in response to their thoughtful and constructive comments.

Writing books is a time-consuming enterprise and best undertaken with the minimum of distraction. For this reason, I am very grateful to the ESRC who supported this work with the award of an extended research fellowship (RES-000-27-0184), thus freeing me from all normal university duties.

Jonathan Evans
Plymouth, March 2007

CHAPTER ONE

Introduction and theoretical framework

It is evident that the human species is highly intelligent and well adapted. Some of our intelligence we clearly share with many other animals; we have well-developed visual and other perceptual systems, complex motor skills and the ability to learn in many ways to adapt to the environment around us. We also seem to be smart in ways that other creatures are not: we have a language system that is complex and sophisticated in its ability both to represent knowledge and to communicate with other humans; we study and attempt to understand a multitude of subjects including our own history and that of the universe; we have devised systems of mathematics and logic; we design and build a huge range of structures and artifacts; we have constructed and mostly live our lives within highly complex economic and social structures. All of these distinctively human things imply an extraordinary ability to reason, entertain hypotheses and make decisions based upon complex mental simulations of future possibilities. I will use the term “hypothetical thinking” as a catch-all phrase for thought of this kind.

It is equally apparent that evidence of human error and fallibility surrounds us. The world is plagued by wars, famines and diseases that in many cases appear preventable. Stock markets collapse under panic selling when each individual acts to bring about the outcome that none of them wants. Doctors sometimes make disastrous misjudgements that result in the disability or death of their patients. Experts often fail to agree with each other and may be shown in hindsight to have made judgements that were both mistaken and overconfident. At the present time, governments of the world are well

informed about the likely progress of global warming and its consequences but seem to be making minimal progress in doing anything to prevent it. Criminal courts continue to convict the innocent and acquit the guilty, with alarming regularity. And so on, and so forth.

It seems vital that psychologists should be able to provide understanding of the mental processes of reasoning and judgements that underlie the actions and decisions that people take. A fundamental premise of the current book is that there are two distinct kinds of thought, which for the moment I will call intuitive and deliberative. Many of our everyday decisions are made rapidly and intuitively because they just feel right. Others are made much more slowly, involving conscious deliberative thinking. Sometimes we have no time for deliberative thought and just have to react quickly to some situation. In fact, the great bulk of our everyday cognitive processing is carried out rapidly and implicitly without conscious thought. Such processes enable us to accomplish a multitude of necessary tasks, as, for example, when we recognize a face, extract the meaning from a sentence, keep our car safely on the road when driving to work (and thinking consciously about something quite different) or attend to the voice of one person in a room containing the babble of many conversations.

Much of our judgement and decision making takes place at this level also. A lot of our behaviour is habitual, so we are not conscious of choosing our direction at a junction on a familiar drive to work. However, something very different happens when we drive to a new location in an unfamiliar town, following verbal directions or trying to read a map. Now we have to engage conscious and deliberative thinking and reasoning to work out the route, identify landmarks, turn at the correct places and so on. In general, novel problems require much more deliberative thought than do familiar ones. When we have to do this kind of thinking it takes time, it requires effort and it prevents us from thinking about other things. Conscious, deliberative thinking is a singular resource that can only be applied to one task at a time. This is one reason that we allocate this kind of thought to tasks and decisions that have great importance for us and make snap intuitive decisions about less important things. However, there is no guarantee that thinking about our decisions will necessarily improve them (see Chapter 5).

Folk psychology – the common-sense beliefs that we all hold about our own behaviour and that of our fellow human beings – involves the idea that we are consciously in control of our own behaviour – we think, therefore we do. The opinion polling industry, for example, is built on the common-sense belief that people have conscious reasons for their actions which they can accurately report. Psychological research, however, seriously undermines this idea (Wilson, 2002). Not only is much of our behaviour unconsciously controlled, but many of our introspections provide us with unreliable information about the extent and the ways in which our conscious thinking controls

our actions. Working out the relative influence of intuitive and deliberative thinking and the interaction between the two systems is a complex problem that must be addressed with the methods of experimental psychology. This enterprise lies at the heart of the current book.

Many of the phenomena to be discussed in this book are described as cognitive *biases*. It may appear that the demonstration of bias implies evidence for irrationality, and it is impossible to study these topics without taking some view on whether and in what way people are rational. Cognitive psychology as a whole studies the workings of the mind at a number of levels. Basic cognitive processes (still incredibly complex and sophisticated) form the building blocks for our behaviour and thought. These include such functions as pattern recognition, language comprehension, memory for events and the acquisition of conceptual knowledge about the world around us. None of these topics has generated debate about human rationality. Our visual systems have limited acuity and our memory systems limited capacity, we assume, because that is simply the way our brains are designed: the way they were shaped by evolution to be. The study of higher cognitive processes, on the other hand – thinking, reasoning, decision making and social cognition – has been somewhat obsessed by the notions of bias, error and irrationality. Author after author provide us with evidence of “bad” thinking: illogical reasoning, inconsequential decision making, prejudice and stereotyping in our view of people in the social world. The study of cognitive biases is something of a major industry.

What exactly is a cognitive bias? One definition is that it is systematic (not random) error of some kind. This then begs the question of what is an error. Psychologists have largely answered the second question by reference to normative systems. Thus reasoning is judged by formal logic; judgement under uncertainty by probability theory; choice behaviour by formal decision theory and so on. Some authors go further and claim that people who fail to conform to such normative standards are *irrational*. Most of the biases studied in cognitive psychology have been defined in this way, and yet this notion is today highly controversial. Some authors claim that people’s behaviour only appears biased or irrational because the wrong normative theory is being applied. For example, if standard logic requires that propositions are clearly true or false, then people’s reasoning in an uncertain world might better be assessed by norms based on probability theory (see Oaksford & Chater, 2001).

In fact, we do not necessarily need to invoke normative rationality in order to think about cognitive biases. We have much lower visual acuity than does a bird of prey, but vision researchers do not accuse us of being biased against distant objects. Similarly, memory researchers do not accuse us of irrationality if we cannot remember a phone number more than seven or eight digits in length. Researchers in this area rarely use the term “bias”, but their findings

certainly indicate the constraints and limitations of human information processing. So we could think about biases of thought and judgement also as indicators of the design limitations of the brain. This is an approach that emphasizes what is known as *bounded* rationality (Simon, 1982). According to this view, we are not inherently irrational but we are cognitively constrained in the way we can reason about the world. For example, it may not be possible to calculate the best choice of action in a given situation, so we settle for one that is good enough.

Another concept of cognitive bias is dispositional: for example, people have different styles of thinking that may be related to personality or to culture. A widely cited claim is that Western people have a more analytic style of thinking, while Eastern people are more holistic or intuitive (Nisbett, Peng, Choi, & Norenzayan, 2001). One style is not necessarily better than the other, but each may fare better or worse on different kinds of task. Combining the dispositional and bounded rationality approaches, we might conclude that people's ability to think in particular ways is biased or constrained not only biologically, in the design of our brains, but also culturally. Either or both kinds of explanation might be induced, for example, to account for biases in social cognition. For example, people seem compulsively to employ stereotypes when thinking about people from an "out-group" with whom they do not share social membership (Hinton, 2000). This could reflect some innate form of social intelligence shaped by evolution, learning of cultural norms passed from one generation to the next, or an interaction of the two.

As we shall see in this book, psychologists studying higher cognitive processes have discovered and documented a wide range of biases. In most cases, these biases have been defined as deviation from a normative standard, leading to a debate about whether or not they should be termed irrational. I have discussed the rationality issue in detail elsewhere (Evans & Over, 1996a), and it will not be the main focus of interest in this book. (I will, however, consider the issue in my final chapter.) The study of cognitive biases should be seen as important for two reasons, whether or not they are deemed to provide evidence of irrationality. First, they establish the phenomena that have to be explained. Second, they may have practical implications for reasoning and decision making in the everyday world. Hence, each bias gives rise both to a theoretical question: "Why do people think in this way?" and to a practical question: "How will this bias manifest itself in real-world behaviour and with what consequences?"

As an example, psychologists have accumulated much evidence that people's evaluation of logical arguments is biased by whether or not they believe the conclusions given (Chapter 4). This is regarded as a bias because logical validity depends only on whether a conclusion follows necessarily from some assumptions and not on whether assumptions or conclusion are actually true. I suppose one could try to move directly from this result to its

practical implications without any real theoretical analysis of the cause of the bias. Such an analysis might, however, conclude that human reasoning is automatically contextualized by prior knowledge and belief and that only a strong effort of deliberative conscious reasoning will overcome this. In my view, understanding of the likely practical implications of the bias is greatly assisted by this kind of theorizing.

In this book, I shall be viewing the phenomena discussed within both a broad and a more specific theoretical framework to be introduced later in this chapter. The broad framework, generally known as “dual-process” theory, has been applied to a wide range of cognitive studies, including learning (Reber, 1993), reasoning (Evans & Over, 1996a; Stanovich, 1999), conceptual thinking (Sloman, 1996), decision making (Kahneman & Frederick, 2002) and social cognition (Chaiken & Trope, 1999). Dual-processing approaches assert the existence of two kinds of mental processes corresponding broadly to the idea of intuitive and deliberative thinking and to more general distinctions between implicit and explicit cognitive processes, such as those involved in learning and memory. Within this general framework, I will, however, present a more specific dual-process theory of hypothetical thinking that updates and extends my earlier heuristic–analytic theory of reasoning (Evans, 1989). In support of this theory, I will discuss phenomena that are drawn mostly (but not exclusively) from two separate but related literatures: the psychology of reasoning on the one hand; and the study of judgement and decision making on the other. Before presenting my general and specific theoretical framework, I shall outline the nature of these two fields of study, including the methods and theoretical approaches that have tended to dominate them.

THE PSYCHOLOGY OF REASONING

I ought, perhaps, to start with the distinction between implicit and explicit inference (see also Johnson-Laird, 1983). Any kind of inference involves going beyond the information given and may technically be regarded as deductive or inductive. Inductive inferences add new information, whereas deductive inferences draw out only what was implicit in assumptions or premises. Both deductive and inductive inferences may be either implicit or explicit in terms of cognitive processing. I shall illustrate this with some examples.

Pragmatic inferences are almost always involved in the comprehension of linguistic statements (see Sperber & Wilson, 1995, for discussion of many examples). Because they typically add information from prior knowledge relevant to the context, they are generally inductive as well as implicit. As a result, such inferences are plausible or probable but not logically sound and may turn out to be incorrect. In accordance with the communicative principle of relevance (Sperber & Wilson, 1995), every utterance conveys a guarantee

of its own relevance, and this licenses many pragmatic inferences. Consider the following dialogue between an adult son and his mother:

“I think I am going to be late for work”

“My car keys are in the usual place”

“Thanks, Mum”.

There will be a context behind this exchange that is mutually manifest to both parties. For example, the son usually travels to work but his mother sometimes lets him borrow her car, which takes 15 minutes off the journey. Hence, the first statement is interpreted as a request to borrow the car, and the reply acquiescence to this request. Neither speaker has actually stated that the car is to be borrowed, so the inferencing is clearly implicit. It is also hardly deductive and can be incorrect. Suppose the dialogue actually went like this:

“I think I am going to be late for work”

“My car keys are in the usual place”

“I am going for a drink after work. Can’t you drop me off?”

The son’s reply clearly signals that the mother’s original inference that he wanted to borrow the car was wrong. This is cancelled by the reply with a further implicit inference: the son wishes to drink and will therefore not drive home afterwards. This kind of inferencing occurs all the time in everyday dialogue, but it is not what the psychology of reasoning is (apparently) concerned with, as we shall see shortly. Note that such implicit inferences can be deductive in nature, as in:

“I can’t play golf this weekend; my sister is visiting”

“Surely, she can spare her brother for a few hours?”

By the conventions of relationships, it follows logically that if X (male) has a sister Y, then X is the brother of Y. This inference is included in the riposte above, but it is most unlikely that either party would have required any conscious reasoning to deduce it. Such inferences are also implicit or automatic but cannot normally be cancelled, unless the premise on which they are based is withdrawn.

What is described as the psychology of reasoning should really be known as the psychology of *explicit* reasoning as it has, at least on the face of it, nothing to do with these kinds of conversational inferences. Instead, psychologists in this field have concentrated on giving participants in their experiments verbal statements from which explicit conclusions need to be inferred. Explicit reasoning tasks can in principle be deductive or inductive, but the field has been generally dominated by the former, using what is known as the

“deduction paradigm” (Evans, 2002a). This method involves giving people some premises, asking them to assume that they are true and then asking them to decide whether some conclusions necessarily follow. This method allows people’s reasoning to be assessed against the normative framework of formal logic. For example, people might be presented with a syllogism that has two premises and a conclusion, such as:

- Some of the blue books are geography books 1.1
- None of the large books are geography books
- Therefore, some of the blue books are not large.

The logical question to be asked is, does the conclusion of this argument necessarily follow from its premises? To put it another way, if the premises of the argument are true, must the conclusion be true as well, no matter what else we assume about the state of the world? The above argument is valid in this sense. The first premise establishes that there exists at least one blue geography book. Since none of the large books are geography books there exists at least one blue book that is not large. Hence, some (meaning at least one) of the blue books are not large. Suppose, we reorder the terms of the conclusion:

- Some of the blue books are geography books 1.2
- None of the large books are geography books
- Therefore, some of the large books are not blue.

Now is Argument 1.2 still valid? The answer now is no. The conclusion would be false if all of the large books are blue. Although there is at least one blue geography book (that is not large), it is perfectly possible that all of the large books are blue. The actual state of affairs, for example, might be:

- 10 small blue geography books
- 20 small red geography books
- 30 large blue history books.

Given this collection of books, both premises of both arguments hold: some of the blue books are geography books, and none of the large books are geography books. The conclusion of 1.1 also holds: some of the blue books are not large. However, the conclusion of 1.2 is demonstrably fallacious because all of the large books are blue. What this illustrates is the semantic principle of validity: an argument is valid if there is no counterexample to it. This principle is favoured by psychologists in the mental model tradition (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991), who have built a popular theory of human deductive reasoning around it.

In contrast with conversational inferences, which are automatic and effortless (though not necessarily logically valid), explicit deductive reasoning tasks of this kind are slow and difficult to solve for most people. In fact, psychological experiments on deductive reasoning show that many mistakes occur with ordinary participants (Evans, Newstead, & Byrne, 1993). In particular, people make many fallacies: that is, they declare arguments as valid when their conclusions *could* be true given the premises, but do not need to be true. Hence, many people would indicate if asked, that both 1.1 and 1.2 above are valid arguments. Syllogistic reasoning is also known to be systematically biased by several factors as we will see later in this book.

How do ordinary people engage in deductive reasoning? For many years, the psychology of reasoning was dominated by two apparently contrasting theories. According to a tradition known as mental logic (Braine & O'Brien, 1998a; Rips, 1994) people have a logic built into their minds comprising a set of inference rules. In the mental logic account of reasoning it is assumed that the content of a particular reasoning problem is stripped out so that the underlying abstract logical form is recovered. Reasoning then proceeds like a proof in formal logic, by application of standard inferential rules or schemas. Consider the following argument:

The car is either a Ford or a Mercedes	1.3
If the car was built in the USA then it is not a Mercedes	
The car was built in the USA	
Therefore, the car is a Ford.	

It is quite easy to see that this is a valid argument, but how exactly do we do this? According to the rule theory we first strip out the content, reducing it to an abstract form:

1. Either A or B
 2. If C then not B
 3. C
- Therefore, A.

Reasoning now proceeds as a mental proof, citing inference rules and the assumptions they require:

4. Not B (Modus Ponens, 2, 3)
5. A (disjunction elimination, 1, 4).

According to the rule of Modus Ponens, if we know that “if p then q” and “p” we can infer “q”. Substituting C for p, and not-B for q, we conclude not-B. The other step requires this inferred statement to be combined with the first

premise: “Either A or B”. The rule of disjunction elimination states that if one component of a disjunctive statement is false, then the other must be true. This leads us to the conclusion A, which can be restated as “the car is a Ford”. The theory requires both that people have such logical rules built in to their minds and that they have a set of effective procedures for applying such rules to draw inferences (see Rips, 1994, for a full computational implementation).

According to the rival mental model account, people do not use inference rules. Rather, they construct what I will call (for reasons that will become apparent later) *semantic* mental models. Such models represent possible situations in the world. Consider the first premise of 1.3:

The car is either a Ford or a Mercedes.

This disjunctive form is ambiguous as it sometimes is used in an inclusive sense (both disjuncts are possible) and sometimes is exclusive. Context (which can influence the construction of mental models) indicates here that the disjunction is exclusive as cars have only one manufacturer. Hence, the statement is represented by two models:

Ford	
	Mercedes

The second premise was:

If the car was built in the USA then it is not a Mercedes.

According to the model theory such conditionals are compatible with three possibilities (a highly contentious assertion, as we shall see in Chapter 3):

Built in the USA	not Mercedes
Not built in the USA	Mercedes
Not built in the USA	not Mercedes.

However, people represent only one model initially in a short-hand form:

Built in the USA	not Mercedes
...	

where “...” is a “mental footnote” that there are other possibilities in which the antecedent is false. This makes Modus Ponens a trivial inference, as once it is asserted that the car is built in the USA it is consistent only with the explicit model in which it is not a Mercedes. This inference in turn eliminates the

second model of the disjunctive premises, leaving only the model in which the car is a Ford, hence supporting the conclusion as a valid argument.

The debate between mental logic and mental model theorists has been difficult to resolve on the basis of empirical evidence, even though each side has made strong claims (Evans & Over, 1996a, 1997). However, it is becoming more and more apparent that the two camps share a common agenda – which I term *logicism* – that is rejected by a number of contemporary researchers, including myself. They assume that deduction is the primary mode of inference and that both logical errors and the massive influence of pragmatic factors (as I shall refer to the effects of problem content and context) interfere with this underlying deductive mechanism. Mental logicians, for example, faced with a mass of evidence of nonlogical influences on deductive reasoning tasks have argued that there is no singular mechanism of reasoning, and that mental logic is supplemented by a whole range of mechanisms discussed by other authors, including pragmatic implicature, pragmatic reasoning schemas (both discussed in Chapter 4) and even mental models (Braine & O'Brien, 1998a). Johnson-Laird and his colleagues have from the start tried to build in some principles to explain why reasoning is competent in principle but defective in practice. For example, it is claimed that working memory limits constrain the construction of the multiple mental models needed to avoid fallacious inference, or that beliefs may bias the process of searching for counterexample models (Johnson-Laird & Byrne, 1991).

Later in this chapter, a dual-process theory of hypothetical thinking will be described that is applied to a range of cognitive tasks, including the deduction paradigm. It will be argued that reasoning is by habit and default pragmatic and not deductive and that only conscious reasoning effort induced by special instructions can result in an effort at deduction. “Logicism” is rejected both as a normative standard (the idea that people ought to be reasoning logically) and as a descriptive approach (describing reasoning as deductively competent in principle). This will also imply, as we shall see, a non-normative definition of cognitive bias. If the brain is not designed as a logical reasoning machine, then it cannot be regarded as malfunctioning when reasoning does not conform to a logical standard. Before introducing this framework, however, some introduction is needed to the other main research paradigm with which this book is concerned.

THE PSYCHOLOGY OF JUDGEMENT AND DECISION MAKING

Decision making is of enormous theoretical interest and practical importance. We actually make thousands of decisions every day in the sense that we (or, at an automatic level, our brains) choose one course of action from among alternatives. How many “decisions” in this sense are involved, say, in a

20-minute drive from home to work? Traffic and weather conditions are never identical even in this most routine and boring of tasks. Most of these decisions (shifting gears, adjusting steering, modifying speed by use of brakes or accelerator) are automatic, but some may be conscious – for example, deciding whether to stop at traffic lights that annoyingly start to change just as we approach them.

Just as the psychology of reasoning has focused on formal and explicit reasoning problems, rather than the huge number of automatic conversational and pragmatic inferences that we make every day, so the academic study of decision making has tended to focus on formally defined decisions of a particular kind. These decision problems tend to be explicit and well defined. There is enormous interdisciplinary interest in economic, business and governmental decision making. Psychologists working in this field have, however, tended to focus on cognitive and social psychological accounts of individuals engaged in decision processes (Hastie, 2001; Koehler & Harvey, 2004). Typically, psychological experiments in this area consist of presenting people with hypothetical scenarios in which they are required to make choices between proposed alternatives, often imagining themselves to be in a role or situation described to them.

Like the psychology of reasoning, the psychology of decision making has a normative theory and a rationality debate. The normative theory, originally introduced from economics by Ward Edwards (1961), is that of expected utility. Most real-world decisions (certainly most that psychologists are interested in studying) are *risky*. They involve consideration of uncertain prospects. The central normative principle of economic decision theory is that people should estimate both the probability (p) of various outcomes and the utility (U , subjective value) to them. They should then calculate the expected utility (EU) of the n outcomes of a decision as follows:

$$EU = \sum_{i=1}^n p_i U_i \quad 1.4$$

Finally, people should act so as to maximize expected utility – that is, choose the action (or inaction) that has the greatest expected utility among the choices present. This model of decision making assumes that people are essentially selfish and that it is optimal to maximize the gain (or minimize the loss) to yourself.

As with logic, it is apparently easy to demonstrate violations of normative decision theory. Consider the case of buying a ticket for the national lottery or buying insurance against your house burning down. From the viewpoint of objective probabilities and monetary values each of these everyday activities involves an expected loss. The lottery collects more in stakes than it gives out

in prizes; similarly the insurance company collects more in premiums than it pays out on losses. Their expected gain is an expected loss to the customer, so why do people persist in these behaviours? Actually, there are many explanations that can be given, showing just how weak the normative theory of decision making really is (see Evans & Over, 1996a, chap. 2). For example, we can argue that people buy lottery tickets because they overestimate the chance of winning (subjective probability) or because their utility function for money undervalues the stake relative to the winnings. Or it can be argued that they have a utility for gambling based on its intrinsic pleasures and so on.

Decision theory even allows us another decision rule (maximin) in which we can prefer an option that has the better security level (least bad outcome). Hence, we buy fire insurance to avoid the worst outcome of losing our house even though it gives us an expected loss financially. Then there is the problem of how far ahead we project the consequences of a decision. Is it rational for a school leaver to choose a university course on the grounds of: (a) how much she will enjoy the course, (b) the first job that it will enable her to get or (c) her career and salary prospects in twenty years' time? Each analysis could lead to widely differing expected utility calculations. All of this means that establishing what constitutes a *bias* in decision making by reference to such an elastic and subjective normative framework is far from easy.

From the viewpoint of hypothetical thinking theory, however, decision making is very interesting. As already indicated, many decisions are made automatically, perhaps in response to past learning. *Consequential* decision making, however, requires hypothetical thinking about future events. We need somehow to imagine the world (in relevant respects) as it might be following a particular choice or action under our control and decide how much we would like to be living in it. Moreover, we need to conduct a set of thought experiments for each possible action and compare their evaluations. Alleged biases in decision making will be discussed in Chapter 5.

An important aspect of the mental simulations required for decision making is the assessment of probability and uncertainty, so much so that the study of judgement under uncertainty has become a large psychological field of study in its own right (Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982). The dominant paradigm has become known as “heuristics and biases”, originally introduced by Amos Tversky and Danny Kahneman (1974) over thirty years ago. In this case, the normative theory that applies is probability theory, with deviations from its prescriptions being regarded as biases. While biases are observed behaviours, heuristics are theoretical constructs. Two of the most famous of these are *representativeness* (Kahneman & Tversky, 1972; Teigen, 2004) and *availability* (Reber, 2004; Tversky & Kahneman, 1973).

The availability heuristic applies when we try to estimate the likelihood or frequency of some particular event. The claim is that we do this by calling to

mind examples of the event. The more easily we can generate such examples, the more frequent we judge the event to be. While this seems reasonable, it can easily be biased, for example by media coverage. Hence, tourists might be deterred from visiting a city due to well-publicized acts of terrorism but not by road traffic accidents, which are a much more probable cause of death and injury that receive little media attention. The representativeness heuristic is applied to judging the probability of a sample given a population or an event given a hypothesis and is based on similarity. Given a thumbnail description of John, for example, we might judge him likely to be an engineer if the description fits our stereotype. However, this could lead to a bias if we ignored the base rate frequency of engineers in the population we are considering. Examples of biases in this literature will be discussed in Chapter 6.

There is another tradition within the psychology of judgement and decision making known as *social judgement theory* or SJT for short. This derives from the psychology of Egon Brunswick who put great stress on the interaction between people and the environment (for a special journal issue on this topic, see Doherty, 1996). Research in this area normally involves multicue judgement when people have to make a single holistic judgement in response to a number of potentially relevant cues. For example, a doctor making a diagnostic judgement might have to take into account a number of pieces of information such as patient symptoms, clinical interview, medical history, demographic variables such as age, gender and occupation, results of diagnostic tests and so on. Of course, some of these cues may be more diagnostic than others.

SJT uses a methodology known as the “lens model” (Cooksey, 1996) in which multiple regression analysis is used to assess the relationship between available cues to the criterion that is being judged on one side of the lens, and the judgements made by individual people on the other side of the lens. This is a clever technique about which I will have more to say later in the book. From the viewpoint of research on biases, it is a powerful method since it allows us to distinguish three different explanations of why people fail to make accurate judgements about the world. It could be that the world lacks predictability or that the judgements lack consistency. If neither of these things is true, but judgements are still poor, then it must be the case that there is a mismatch between the judge’s model and the “world’s” model. For example, if a personnel manager consistently prefers to select young males for positions in which neither age nor gender are relevant to performance, this will reduce his performance. Bias of this kind is easily detected with the methods of SJT.

DUAL-PROCESS THEORIES OF THINKING

The general framework for considering explanations of cognitive biases in this book is that of dual-process or dual-system theories. As mentioned earlier,