1

Introduction

In this opening chapter I shall introduce my core claim, provide an over-view of the chapters to follow, and make some remarks about the aims and scope of the project.

I THE CORE CLAIM

The concept of belief is a multi-faceted one. A belief ascription may pick out an episodic thought or a long-held opinion, a considered conviction or an unthinking assumption, a deliberate judgement or a perceptual impres-sion. In the first person, it may express a tentative suggestion or an item of profound faith, a speculative hypothesis or a confident assertion, a rou-tine recollection or a revelatory self-insight. This diversity is not in itself a problem; many everyday concepts have a similar richness of structure. The concept of belief is special, however. For many philosophers and psycholo-gists believe that it can be co-opted to play a very precise role. They believe that our everyday practices of psychological description, explanation, and prediction – practices often referred to as *folk psychology* – are underpinned by a primitive but essentially sound theory of human cognition, whose concepts and principles will be central to a developed science of the mind. That is to say, they believe that the concept of belief, together with those of other folk-psychological states, can be integrated into science and applied to the business of serious scientific taxonomy. I shall refer to this view as *integrationism*.[1]

In a weak form, at least, integrationism is an attractive position: there is a strong case for thinking that the broad explanatory framework of folk psychology is as sound as those of other special sciences (Fodor 1987,

---

[1] The term 'folk psychology' is often used to refer to the putative theory underpinning our everyday practices of psychological explanation and prediction, as well as to the practices themselves. To avoid confusion, I shall use the term only in its broader sense, to refer to the practices.

ch. 1; Botterill and Carruthers 1999, ch. 2). Of course, folk psychology
cannot be integrated into science just as it stands. At the very least, it will
be necessary to identify its *theoretical core* – to analyse its central concepts
and to articulate its fundamental principles and assumptions. It may also
be necessary to refine this core and make it more precise, and perhaps
even to revise it in some ways. And, of course, we shall need to con-
firm that the resulting theory is sound and, more specifically, compatible
with what we know about human biology and neurology. A huge amount
of work has been devoted to these tasks, yet no consensus has emerged.
There are deep and seemingly intractable disputes about the nature of
belief – its metaphysics, semantics, causal role, and relation to language.
And there are continuing worries about the compatibility of the folk the-
ory with our best neuroscientific theories. In the light of these problems,
some writers have concluded that we should abandon integrationism and
eliminate folk-psychological concepts from science, while others argue
that only attenuated versions of the folk concepts can be retained (for the
former view, see Churchland 1979, 1981; Stich 1983; for the latter, Clark
1993b; Dennett 1987; Horgan and Graham 1990).

In this book I shall be outlining an alternative integrationist strategy
which promises to resolve some of the theoretical disputes just mentioned
and to establish the compatibility of folk psychology and neuroscience,
while at the same time preserving a robust common-sense conception of
the mind. The strategy involves questioning an assumption common to
most existing integrationist projects. This is that belief is a *unitary* psy-
chological kind – that whenever we ascribe a belief to a person, creature,
or system, we ascribe essentially the same *kind* of state.[2] Of course, no
one denies that belief has varied aspects and manifestations – it is widely
accepted that beliefs can be both *occurrent* and *standing-state*, *explicit* and *tacit*,
conscious and *non-conscious*, and so on. But it is generally assumed that these
are different aspects or variants of the same underlying state. So occur-
rent beliefs can be thought of as activations of standing-state beliefs, tacit
beliefs as dispositions to form explicit beliefs, conscious beliefs as beliefs
that are the object of higher-order beliefs, and so on. This assumption –
the *unity of belief assumption*, as I shall call it – shapes the direction of most
integrationist projects. Typically, theorists begin by articulating a *core notion*
of belief, and then go on to show how different varieties of belief can be

---

[2] Here, as throughout, I focus primarily on belief. I assume, however, that parallel claims can
be made for desire, and perhaps for other mental states, too, and shall occasionally indicate
how these would go.

Introduction

defined in terms of it. The unity of belief assumption is often coupled with a parallel assumption about reasoning. Theorists tend to assume that this, too, has a uniform character, and to advocate single-strand theories of reasoning, which identify thought processes with a single, generic kind of activity – computational operations in a mental language, say, or asso-ciative processes of some kind. This assumption – I shall call it the *unity of processing assumption* – is not quite so pervasive as the parallel one about belief, and has occasionally come under challenge from psychologists. It is common in the philosophical literature, however, and shapes many of the debates there.

There have, it is true, been dissenting voices, suggesting that the apparent uniformity of folk-psychological discourse masks important psychological distinctions. Some writers distinguish passive belief formation from active judgement. Philosophers of science, too, commonly mark a distinction between partial and flat-out belief (sometimes called 'acceptance'). And Daniel Dennett has argued that we must distinguish non-verbal beliefs from a class of language-involving cognitive states which he calls *opinions*. The distinction between the two states is, he claims, a very important one:

> My hunch is that a proper cognitive psychology is going to have to make a sharp distinction between beliefs and opinions, that the psychology of opinions is really going to be rather different from the psychology of beliefs, and that the sorts of architecture that will do very well by, say, nonlinguistic perceptual beliefs (you might say animal beliefs) is going to have to be supplemented rather substantially in order to handle opinions. (1991b, p. 26)

Indeed, Dennett suggests that a failure to distinguish these states lies at the root of many philosophical misconceptions about belief (see the references to 'opinion' in Dennett 1987; see also his 1991d, p. 143, and 1994, p. 241).

It would not be too surprising if something like this were true. After all, everyday users of folk psychology are interested primarily in behavioural prediction and explanation, not precise psychological taxonomy. If two psychological states or processes were similar enough to be lumped together for everyday purposes, then we should not expect folk psychology to make a sharp distinction between them – though it might register their distinctness in indirect ways. The states and processes in question might nonetheless differ significantly, and it might be important for a developed psychology to distinguish them – even if we continued to conflate them for everyday purposes. That is to say, folk-psychological concepts may

turn out to be what Block calls *mongrel concepts* (Block 1995), and it may
be necessary to distinguish different versions of them if they are to be
integrated into science. The introduction of new distinctions like this is
common in integrationist projects. Consider, for example, how psycho-
logical theory has adopted the common-sense concept of *memory*, while
at the same time distinguishing various kinds of it – long-term, short-
term, episodic, procedural, semantic – each with different functions and
properties.

I believe that something similar will happen with the folk concepts of
belief and reasoning. These, I shall argue, conflate two different types of
mental state and two different kinds of mental processing, which form two
distinct levels of cognition. That is, I shall be arguing that the search for
a single theoretical core to folk psychology is misguided: folk psychology
has – or a rational reconstruction of it will have – two distinct theoretical
cores. (I shall say more in a moment about the relative roles of analysis
and rational reconstruction in this project.) In short, we need a *two-strand*
theory of mind. Only by developing such a theory, I believe, can we
resolve some deep disputes about the mind and provide a sound basis for
integrating folk psychology into science.

2 AN OVERVIEW OF THE BOOK

To date, there have been few sustained attempts to develop two-strand the-
ories of belief. Few theorists have sought to link up the various distinctions
that have been proposed or to explore their implications for issues in phi-
losophy of mind. Dennett is one of the few exceptions here, drawing on
a number of sources in a richly suggestive essay (Dennett 1978a, ch. 16).
However, he does not work out his ideas in a systematic way and tends to
treat opinion as something of a cognitive side-show, which is not directly
implicated in reasoning and the guidance of action. And while some psy-
chologists have advanced 'dual-process' theories of reasoning, there have
been few attempts to integrate these theories with two-strand theories of
belief or to consider their philosophical consequences.

This book aims to remedy these omissions. Chapter 2 begins by high-
lighting some divisions in the folk notion of belief – divisions relat-
ing to consciousness, activation level, degree, method of formation, and
relation to language. These divisions, I argue, are real and run deep and
link together in a natural way to yield a tentative two-strand theory of
belief – the first strand non-conscious, partial, passive, and non-verbal, the

4

second conscious, flat-out, active, and often language-involving. I then
move on to look at similar divisions in our view of reasoning. Again, I
argue that these run deep and indicate the need for a two-strand theory –
the strands corresponding closely to the two strands of belief. The final
section of the chapter looks at some further divisions in folk psychology,
concerning the ontological status of belief and the function of psycholog-
ical explanation. I identify two broad interpretations of folk psychology,
which I call *austere* and *rich*, and which correspond roughly to the views
of philosophical behaviourists and functionalists respectively. On the aus-
tere interpretation, folk psychology is a shallow theory, which picks out
behavioural dispositions and offers explanations that are causal only in a
weak sense. On the rich interpretation, it is a deep theory, which aims to
identify functional sub-states of the cognitive system and to offer causal
explanations of a more robust kind. I suggest that these two interpretations
each have a firm basis in the folk outlook and that a reconstructed folk
psychology needs to admit both. The two *interpretations*, I argue, indicate
the need for two *theories*, corresponding to the two strands of mentality
identified earlier: an austere theory for the non-conscious strand, and a
rich theory for the conscious one.

It is one thing to identify two strands of mentality, of course, another
to construct a substantive two-strand theory of mind. A developed theory
will need to explain how the two strands are related to each other, what
role they play in reasoning and action, and how they combine to form a
single intentional agent. Chapters 3, 4, and 5 are devoted to this task.

Chapter 3 begins by setting out some challenges to the proposed
two-strand theory. Prominent among these is what I call the 'Bayesian
challenge' – the challenge of reconciling our common-sense belief in the
existence and efficacy of flat-out belief with a Bayesian view of rational
decision-making. I then review some precedents for a two-strand theory
of mind, seeking hints as to how to develop the theory and respond to the
challenges. I focus in particular on possible models for the conscious, flat-
out, language-involving strand of belief, and on suggestions as to how this
strand might be related to the other, non-conscious strand. Although none
of the models examined fits the bill exactly, I identify several promising
ideas, including the *behavioural view* of flat-out belief developed by some
Bayesians, Dennett's picture of the conscious mind as a *virtual machine*, and
Cohen's account of acceptance as a *premising policy*. I conclude the chapter
by suggesting how elements of these views can be combined to give a pic-
ture of the conscious mind as a *premising machine*, formed by the adoption

and execution of premising policies, and driven by non-conscious, partial beliefs and desires.

Chapter 4 is devoted to filling in the picture of the premising machine sketched in the previous chapter. I discuss the nature and scope of premising policies and distinguish several varieties of them, including a goal-oriented form. I then look at what is involved in executing these policies and what role natural language plays in the process. Finally, I consider how premising policies are related to other mental states and how they influence action. Crucially, I argue that an agent's premising policies are *realized in* their non-conscious, partial beliefs and desires – and thus that the premising machine constitutes a distinct level of mentality which supervenes on the one below it.[3] To emphasize the point, I call premising policies *supermental* states, and the level of mentality they constitute the *supermind*. By analogy, I call the non-conscious attitudes in which the supermind is realized the *basic mind*. Because the supermind is realized in the basic mind, I argue, supermental explanations of action are not in competition with those pitched at the basic level. Rather, each corresponds to a different level of organization within the agent.

Chapter 5 shows how we can use the framework developed in the previous chapter to flesh out the two-strand theory outlined in chapter 2. I begin by arguing that conscious, flat-out beliefs can be identified with a particular subclass of premising policies, and thus that they, too, are supermental states. The upshot of this is that our two-strand theory of mind becomes a *two-level* one, with conscious, flat-out states realized in non-conscious, partial ones. (In line with the terminology adopted earlier, I call the former *superbeliefs* and the latter *basic beliefs*.) I then go on to highlight the attractions of this view, and to show how it can resolve the challenges posed in chapter 3. The chapter concludes with some remarks on the function of the supermind. I argue that supermental capacities carry with them considerable cognitive benefits. The supermind is a slow but highly flexible system, which can kick in whenever faster but less flexible basic processes fail to yield a solution. Moreover, because supermental processes are under personal control, we can reflect on them, refine them, and supplement them. The flexibility, adaptability, and improvability of human cognition flow directly from the supermind.

---

[3] Throughout this book I use 'they' as a gender-neutral third-person singular pronoun. This usage has a long history in English prose and is, I think, the least inelegant way of avoiding an impression of gender bias.

Introduction

Having developed the core theory and shown how it can vindicate some important aspects of folk psychology, I then move on to consider some further folk commitments and to show how these, too, can be vindicated by the theory.

There is a case for thinking that folk psychology makes some substantial assumptions about the functional architecture of the cognitive system. In particular, it has been argued that there is a folk commitment to the theses of *propositional modularity* and *conceptual modularity* – accounts of how propositional attitudes and their component concepts are stored and processed (Ramsey et al. 1990; Davies 1991). The idea that there is a folk commitment to these theses is worrying, since they in turn seem to involve claims about the architecture of the brain, and therefore to run a risk of empirical falsification. If the folk are committed to them, then their conception of the mind may be seriously mistaken and ripe for revision or even elimination.

In chapters 6 and 7 I show how the two-level theory developed in earlier chapters can vindicate folk psychology's architectural commitments. Chapter 6 deals with propositional modularity and chapter 7 with conceptual modularity. In each case I begin by arguing that there is indeed a folk commitment to the thesis, building on arguments in the literature. I then show that this commitment can be vindicated at the supermental level, without involving claims about the structure of the brain. I show that the supermind exhibits both propositional modularity and conceptual modularity, and thus that the folk assumptions are correct. I argue, however, that there is no inference from this to claims about the structure of the brain. The supermind is implemented, not in the hardware of the brain, but in basic-level intentional states and actions. And the basic mind need not itself exhibit propositional or conceptual modularity in order to support a supermind that does. The upshot of this is that the folk architectural commitments are compatible with any account of the underlying neural architecture. Given this, the threat to folk psychology – in this guise at least – vanishes.

The final chapter outlines some further applications of the proposed theory – starting with a discussion of akrasia and self-deception. These conditions can seem puzzling, and it is sometimes suggested that they reveal the presence of conflicting subagents within the human psyche (Davidson 1982; Pears 1984). Here a two-level theory offers a different and, I think, more attractive perspective. The conditions can be thought of as involving a conflict, not between subagents, but between levels of

mentality – the attitudes at one level in tension with those at the other. In each case, I sketch the two-level account and show that it offers an economical way of resolving the associated puzzles. The chapter then moves on to look at first-person authority. Again, this can appear puzzling. How can we be authoritative about our mental states, given that there are independent behavioural criteria for their possession (Heal 1994)? And, again, the present theory offers a fresh perspective. The thought is that first-person authority proper extends only to supermental states, and is primarily a matter of control. Superbeliefs can be actively formed and processed, and in self-ascribing these states we are not simply *reporting* that we meet the criteria for their possession, but *committing* or *recommitting* ourselves to meeting them. The authority attaching to such an ascription, then, is that of a sincere commitment, rather than that of a reliable report.

In addition to helping to clarify philosophical debates about belief, the theory developed here may have application to issues in developmental and clinical psychology, and the final chapter closes with some brief remarks on this. In particular, I suggest that the theory may be able to shed light on the nature of autism.

3 METHODOLOGICAL REMARKS

It may be useful at this stage to add some remarks about the scope and status of the theory I shall be developing.

First, note that in distinguishing different kinds of belief I shall focus on differences in structure, function, and constitution, rather than content. This is not because I think that the two do not differ in their representational properties. We may need a different theory of content for each kind of belief, and there may also be differences in the range and determinacy of the contents associated with each. But it makes sense to consider broad structural and constitutive questions first. The distinction between the two kinds of belief can be drawn most clearly in this way, and once it is in place, questions about content may become more tractable. The same semantic questions will arise for each type of belief, and we shall be in a better position to address them when we know what kind of attitude is involved in each case. I shall say something about concept possession at the two levels in chapter 7, but there will be no space for extended discussion of issues of content. In this respect the present book can be thought of as preparing the ground for future work.

8

Introduction

Secondly, in proposing a two-level theory of mind, I do not mean to claim that there are *only* two levels of cognition. My aim is to reconstruct *folk* psychology – to show how we can bring our various beliefs about the mind into a coherent theory – not to provide a complete framework for psychological theorizing. In particular, I do not mean to deny the existence of a level of sub-personal psychology underlying the folk-psychological levels (we might call it a 'sub-mind'), though I do claim that the folk are uncommitted as to the existence and nature of such a level. I shall say more about this later.

Thirdly, I want to add a note about the account of the basic mind I shall be defending. As I mentioned, I shall advocate what I call an *austere* view of this level, which treats mental states simply as behavioural dispositions. Folk psychology's architectural commitments, I shall argue, relate to the supermind, not the basic mind. (Again, this does not mean denying the existence of sub-personal psychological structures underlying the disposi-tions that constitute the basic mind – denying, for example, that there is a sub-personal language of thought or a range of domain-specific cognitive modules. The claim is merely that *folk psychology* does not postulate such structures.) This is, I think, the correct view to take. As I shall argue in chapter 2, there is a strand of folk psychology which has no architectural commitments, and one of the virtues of a two-level theory is that it can reconcile the existence of this strand with that of another which does have such commitments. Moreover, it is tactically the right position for me to adopt. For one of my aims is to show how a richly structured supermind can be realized in basic mental states and processes. And in adopting an austere view of the basic mind, I present myself with the hardest case here: if I can show that a richly structured supermind can be realized in an aus-tere basic mind, then it should not be *more* difficult to show that it can be realized in a richer one. However, nothing in my account of the supermind *relies* on an austere view of the basic mind, and it is possible to endorse the former while rejecting the latter. So if you balk at austerity, then feel free to substitute whatever view of the basic mind you prefer. In doing so, you will deprive the account of some of its conciliatory power, but the overall shape of the two-level theory and the description of the nature and function of the supermind will remain unaffected by the change.

Finally, let me say something about the status of the two-level theory I shall develop and the means by which it will be derived. The first part of this is easily done. The theory is an empirical one – a model of how the human mind might be organized – and its evaluation will require empirical

investigation. The theory will not be *derived* by empirical investigation,
however, and I shall not be drawing on experimental data in support of
it – though I shall aim to say nothing that is incompatible with it.[4] The
exclusion of such data reflects the scope of the present inquiry. If folk
psychology is to be integrated into science, then two things must be done.
It will be necessary, first, to identify and regularize its theoretical core,
and secondly, to establish that the resulting theory is true. The primary
aim of this work is to accomplish the first task – to articulate a theory
which best systematizes our common-sense intuitions about what the
mind is like. My theorizing will thus be constrained by a belief in the
fundamental soundness of the descriptive and explanatory practices of
folk psychology, and the data upon which I shall draw will be the product,
not of experiment, but of reflection on those practices. This is not to say,
however, that the work will be merely one of conceptual analysis. I do not
claim that a two-level framework is revealed simply by careful examination
of folk-psychological practice. I do not think that the practice is sufficiently
well-defined for that. Nor do I claim that a two-level framework is tacitly
assumed by users of folk psychology. I suspect that, for the most part, they
simply conflate the states and processes that I aim to distinguish. (The gen-
eralizations which underpin everyday psychological discourse are, I think,
sufficiently loose to hold true of both levels of cognition.) So while the
present work will begin with conceptual analysis, highlighting a number
of divisions within folk psychology, it will not stop there, but will go on
to engage in rational reconstruction – to seek to identify the theoretical
framework which best regiments our folk usage. Thus the aim will be,
not to elucidate a pre-existing theoretical framework, but to supply one
where it is lacking. And the resulting theory will be the product, not only
of an analysis *of* folk discourse, but also of abductive inference *from* it.

This is not all, however. The theory will also have important implica-
tions for the second of the two parts of the integrationist project – the task
of showing that the core folk theory is true. For if the arguments in the
later chapters are sound, then this task will take on a very different aspect.
According to those chapters, folk psychology involves no commitments as
to the internal architecture of the brain: the supermind is constituted by

---

[4] This is a self-denying ordinance, since there is much experimental evidence that could be
cited in support of the thesis. There is a particularly interesting consilience between the
ideas that will be developed here and the *dual-process* theories of reasoning developed by
Jonathan Evans and David Over, among others (see Evans and Over 1996). I shall say a
little about these theories in chapter 8.