

Rationality
+
Consciousness
=
Free Will

David Hodgson

OXFORD
UNIVERSITY PRESS

Contents

Preface ix

Introduction 3

The problem of free will and responsibility 3

Compatibilism 4

Haldane's argument 5

How I will proceed 6

1. Foundational Beliefs 10

Can I be certain that I exist? 10

The need for language 14

Experiences and the external world 15

Foundational beliefs 18

2. Truth and Rationality 20

Truth 20

Relativity of truth? 24

Rationality 26

Fallacies and biases 29

Stich's argument 32

A legal example 34

Core assertions about truth and rationality 36

3. Plausible Reasoning 37

Formal and informal reasoning 37

Induction 40

Bayes' Theorem 43

Illustration of Bayes' Theorem 45

Levels of cognitive processes 48

Core assertions about plausible reasoning 53

4. Consciousness and Decision-Making 55

Dual aspects 55

Characteristic features of conscious experiences	59
Subjectivity	60
Qualia and unity	62
Neural correlates of consciousness	66
The efficacy of conscious experiences	68
Three questions	72
Rule-determined processes do not need consciousness	74
Core assertions about consciousness and decision-making	77
5. Gestalts and Rules	79
The argument outlined	79
Laws and rules	85
The Game of Life and computation	89
Tricks of consciousness	92
Some further thoughts	95
Core assertions about gestalts and rules	97
6. How Gestalts Promote Rationality	98
Evolutionary origins	99
Aesthetic judgments	102
Plausible reasoning	108
Conclusion	111
Core assertions about how gestalts promote rationality	113
7. Science and Determinism	114
A lawful universe	115
Quantum mechanics	118
The free will theorem	121
Explanation of the theorem	124
Implications	126
Time and the block universe	128
Core assertions about science and determinism	132
8. Neuroscience and Conscious Choice	133
Science and the brain	133
A general picture	136
The Cartesian theatre	138
The scale and nature of quantum effects	140
Libet, Gazzaniga and Wegner	148
Core assertions about neuroscience and conscious choice	151

9. Indeterministic Free Will	153
Will and responsibility	154
Comparison with Kane	159
Agent-causation	162
Compatibilism	166
Assessment of compatibilism	171
Does luck swallow everything?	173
More about luck	177
Core assertions about indeterministic free will	179
10. Value Judgments	180
A different philosophical approach	181
Natural imperatives	185
Absolute imperatives	187
Prima facie imperatives	190
No reasonable irreconcilable differences	192
Why be moral?	196
Good, evil and beauty	197
Community practices and laws	198
Legal systems	201
Capacity for reasonable value judgments	203
Core assertions about value judgments	204
11. Responsibility and Retribution	206
Responses to wrong conduct	206
Overview	210
Australian criminal law	213
Retribution as a restriction on State compulsion	216
Why retribution should be maintained	218
Philosophical bases for retribution	223
The future of retribution	225
Core assertions about responsibility and retribution	228
12. The Big Picture	230
The scientific account	230
An experienced universe	231
Constraint, empowerment and guidance	234
Religious belief: a subject for rational enquiry	236
A value-embedded universe	237
Where do we come from?	239

Where are we going?	240
Can more specific beliefs be supported?	243
Potential for evil and good	244
Core assertions about the big picture	246

Appendix A

Why Bayes' Theorem Works	249
---------------------------------	------------

Appendix B

Against Fundamentalism: Biblical Morality	251
--	------------

Abraham and Isaac	251
-------------------	-----

The Passover	252
--------------	-----

The Promised Land	253
-------------------	-----

The New Testament	253
-------------------	-----

References	255
-------------------	------------

Index	261
--------------	------------

Introduction

THE PROBLEM OF FREE WILL AND RESPONSIBILITY

In *Red*, the final film of Krzysztof Kieslowski's *Three Colours* trilogy, neighbours throw a stone through a retired judge's window, and the judge muses, 'I wonder what I'd do in their place.' After a pause, he says, 'The same thing.' The girl who's with him asks, 'You'd throw stones?' The judge replies, 'In their place, of course. And that goes for everyone I judged. Given their lives, I would steal, I'd kill, I'd lie. Of course I would.' And he explains why his life took the course it did: 'All that because I wasn't in their shoes, but mine.'

Is it the case, as this exchange suggests, that everything we do is the inexorable playing out of the role in life given to us by our genes and our circumstances, by nature and nurture? Or is it the case that we have real alternatives and real choices about what we do, so that our actions and our lives are not entirely determined by the luck of our genes and circumstances, but are influenced by decisions we can freely make and for which we are truly responsible?

The problem of free will and responsibility is as old as philosophy itself, and has been the subject of countless books and articles. There are those who think this problem is now resolved, because science shows that we human beings are physical systems that change over time in accordance with the laws of nature that govern all physical systems, and thus cannot do other than as determined by those laws. According to these views, whatever we do is either the inevitable product of our genes and our circumstances, or of chance occurrences, or of some combination of the two; and there can be nothing in our genes or circumstances (or indeed in chance or the laws of nature) for which we can be ultimately responsible. Thus, everything flows from things that happened before we were born or are otherwise outside our control, so that we cannot have free will in any meaningful sense and cannot be truly responsible for our decisions and actions.

Some who think along these lines claim that these views provide a sympathetic account of human conduct, which humanely avoids the judgmental approach of blaming people for their wrong actions (which, on these views, they can't help doing). They say in particular that criminal conduct should properly be regarded as an illness to be

treated rather than wrongdoing to be punished, and that any idea of retribution is based on primitive urges for revenge and should be abandoned.

I disagree with these views. In this book I set out a systematic and cumulative argument for the view that, while the engagement of our genes and our circumstances with laws of nature makes an enormous contribution to the way we are and the way we behave, we nevertheless do make real decisions that are not themselves wholly determined by that engagement and/or by chance, but are in a fundamental way truly *down to us*. I argue that these choices are influenced by input from our conscious experiences that is not determined by laws or rules of any kind, but rather is the exercise of a capacity that our consciousness gives us; and that we can thereby be partly responsible not only for what we do, but also for the way our own characters turn out.

I will also suggest that the idea of retribution is far from primitive and inhumane, but is rather a foundation of human rights. In short I say that if we do not punish people *because they are guilty*, there is less reason to refrain from punishing people *if and because they are innocent*. If it is regarded as acceptable that government officials treat citizens in any such manner as appears to be most beneficial, irrespective of whether persons so treated have done anything to *deserve* that treatment, the way is left open for practices like putting political dissidents into prisons or mental asylums. Respect for human rights requires that, with limited exceptions, governments refrain from interfering with the freedom of citizens unless the citizens have acted in breach of a publicly stated law, in circumstances where they are *responsible* for the breach and can fairly be regarded as *deserving* punishment—in which case it becomes permissible to impose punishment that is reasonably proportionate to the criminality of the conduct in question.

I also believe that, to make the most of life, one needs to recognise one's own responsibility for one's own conduct and (at least partially) for one's own character. Denial of responsibility does not necessarily result in fatalism, but it is not (I contend) conducive to efficacious aspiration for virtue and achievement, or resolution to seek them in one's own life.

COMPATIBILISM

So far I've written as if there are just the two stark alternatives of (1) views denying free will and responsibility and (2) views, such as that advocated in this book, asserting our capacity to make decisions that are not wholly determined by engagement of our genes and circumstances with laws of nature (or by some combination of such engagement and chance). However, there is also a third class of views, to the effect that *free will is compatible with determinism*: according to these views, it is fully consistent with our having free will and responsibility that our decisions and actions be wholly determined by engagement of our genes and circumstances with laws of nature.

These compatibilist views are advocated by many philosophers, perhaps a majority of contemporary philosophers who have addressed the problem of free will and responsibility; and the question of whether or not free will is compatible with determinism has given rise to extensive and sophisticated philosophical debates. I will not be entering into the detail of these debates, because of what I see as the strength of considerations in favour of the account of indeterministic free will and responsibility given in this book.

However, in chapter 9 I will compare my account with compatibilist views, and say a little about the arguments for the position that free will is compatible with determinism.

HALDANE'S ARGUMENT

One of the traditional arguments in favour of free will is the argument attributed to biologist J. B. S. Haldane to the effect that we can't have reasonable grounds for believing determinism to be true, because if determinism were true we would simply believe whatever we were caused to believe by deterministic processes, which govern our beliefs just as they govern everything else.

Two difficulties with this argument have been recognised.

One is that it assumes that reasonable grounds for belief cannot be deterministic, whereas it might be thought that our abilities to perceive accurately and to draw reasonable inferences depend for their reliability upon processes that unfold regularly in accordance with laws of nature and/or rules of logic—that is, on processes that are deterministic. Thus, it can be argued, determinism supports the reasonableness of our beliefs rather than the reverse.

The other is that the argument is purely negative: it doesn't give any positive reason to believe in free will. Certainly, the argument doesn't explain how or why indeterminism or free will could provide any sounder basis than determinism for reasonable belief. It may be contended that to suggest any basis for reasonable belief other than deterministic law-governed processes is to appeal to something mysterious or even magical, a black box of which no explanation can be given.

This book may be regarded in part as a reworking of this traditional argument, in a way that deals head-on with these two difficulties.

As regards the first difficulty, I carefully consider the nature of human rationality, and I argue that essential to this rationality is the ability to engage in *plausible reasoning*; that is, in reasoning in which the premises or data do not entail the conclusions by virtue of applicable rules for good reasoning, but rather support them as a matter of reasonable albeit fallible judgment. And I offer an explanation as to how that kind of reasoning could work in a way that is reasonable yet not determined by rules or laws of any kind.

As regards the second difficulty, I argue that there are very strong grounds for accepting that human beings have the ability to make reasonable decisions about what to do and what to believe (and to deny this would be to make all intellectual endeavours pointless). Careful consideration of this ability shows that it requires *both* consciousness *and also* the capacity to make choices that are reasonable yet not determined by laws or rules—so that we have every reason to believe that we have this very capacity, and thus that we have free will. I also suggest this capacity is no more mysterious than consciousness itself.

My conclusion is that free will, so understood, is *necessary* for the ability that we actually have to make reasonable decisions, and thus is necessary if there is to be any point to intellectual endeavours.

HOW I WILL PROCEED

So in this book I argue that it's reasonable to believe we have free will, in the sense that we make real choices that are not themselves wholly determined by our genes and environment and/or by chance, but are in a fundamental way truly down to us. Of course, scientific discoveries about how the world works and how our brains work are relevant to this view, and some would say they are incompatible with it; and I will have regard to these discoveries. But my argument will take account also of wider considerations.

Basic to my argument are the ideas (1) that in order to arrive at reasonable and justified beliefs we must use the abilities we have that are best able to produce such beliefs and such justification; and (2) that these abilities are captured by the concept of *rationality*, which I understand as not being limited to intellectual ability but as including all those perceptual, intellectual and emotional capabilities by use of which we can (at least sometimes) make reasonable decisions about what to believe and what to do.

In relation to many questions it is rational to rely on the scientific method of formulating and testing hypotheses, in order to arrive at reasonable beliefs about the world; but it is important to appreciate that, although this method is a very important and reliable way to arrive at reasonable beliefs, it is not the only way. For one thing, as I will show, the scientific method itself presupposes the truth of some prior beliefs; and beliefs based on the scientific method cannot be more reliably true than the prior beliefs on which the validity of the scientific method depends. I will argue that these prior beliefs have rational support that does not depend on the scientific method.

Rationality is thus more fundamental than the scientific method; and it is also wider, in that it can be applied to questions that cannot be addressed by the scientific method, such as questions as to the nature of rationality itself, and also moral questions about how we should conduct ourselves and what things we should value.

Coming now to the scheme of the book, in the first three chapters I address questions concerning the rational basis of our beliefs, at a level more fundamental than the scientific method.

I start in chapter 1 by considering what are my own most certain and most fundamental beliefs, and I suggest they are the most reasonable starting points for addressing questions such as these. Like Descartes, I begin by considering what if anything I can be absolutely certain about; but unlike Descartes, I find that I cannot be certain about my own existence, but only that *conscious experiences occur*. As I explain, this is partly because of the need for a *language* with which to express and communicate beliefs of any complexity, and thus the dependence of complex beliefs on prior beliefs about language and language use; and this in turn requires prior beliefs about language users and a world in which language is used.

I begin chapter 2 by saying a little about *truth*. A central aim of my enterprise is to arrive at and justify true beliefs about free will and responsibility, on the way to achieving some understanding of these matters; so it is important to make explicit my conception of what counts as truth. And then I address some general considerations concerning rationality. As I have already suggested, rationality is presupposed in any undertaking to arrive at and justify true beliefs, and it is useful to articulate what is involved in this presupposition.

In chapter 3, I consider one very important aspect of our rationality, namely our ability to engage in plausible reasoning; that is, as mentioned above, reasoning in which the premises or data do not entail the conclusions by virtue of applicable rules, but rather support them as a matter of reasonable albeit fallible judgment. I argue that even the scientific method depends on plausible reasoning, and that plausible reasoning cannot be fully explained in terms of rules for good reasoning.

In the following three chapters, I argue that our consciousness contributes to our rational decision-making in ways that are not wholly determined by rules or laws of any kind, and thus that our decisions can be both rational and not determined by a combination of pre-existing circumstances (including our own characters) and laws of nature. I conclude that we do make real choices that are not themselves wholly determined by our characters and circumstances or by chance, but are in a fundamental way truly down to us.

In the first of these three chapters, chapter 4, I argue that consciousness must have a role to play in our decision-making and that this role is not one that could be performed by rule-determined processes (if it were, consciousness would be a superfluity). I identify the three distinguishing features of conscious processes that I contend contribute to decision-making, namely subjectivity, qualia and unity.

In chapter 5, I offer a specific and straightforward proposal as to what the role of consciousness is, namely to enable an organism to be responsive to circumstances grasped as wholes, not just to their constituent features that engage with laws or rules.

And I support this proposal by an original argument of mine, to the effect that in our conscious experiences we grasp *and respond to* whole feature-rich gestalts, which do not as wholes engage with any applicable laws or rules.

In chapter 6, I consider in more detail how the grasp of gestalts contributes to reasonable decision-making, in particular (1) by non-human animals that have conscious experiences somewhat like ours, (2) in making aesthetic judgments, and (3) in plausible reasoning generally. My conclusion is that indeed our reasonable decision-making utilises conscious experiences in which we grasp and respond to whole feature-rich gestalts in ways that are not determined by laws or rules.

Having reached that point, in the next two chapters I consider how this conclusion stands with science in general and neuroscience in particular.

In chapter 7, I consider whether the general picture of the world suggested by the physical sciences is compatible with my conclusion. I argue that the best contemporary physical theories support two propositions favourable to my position, namely that causation is not local and that the world is not deterministic. I also argue that the so-called block universe view, denying reality to the passage of time and suggesting an unchanging (and in that sense deterministic) universe, is not reasonably tenable.

In chapter 8, I give an account of how the brain could work consistently with my approach, and consider whether this is compatible with what neuroscience tells us about the brain. To the extent that neuroscience seems to leave no room for any contribution from conscious experiences over and above their neural correlates, I contend this is a manifestation of current neuroscience's inability to give a satisfactory account of consciousness generally, and does not count strongly against my approach.

In the next three chapters I discuss how my views on conscious decision-making support ideas about free will, responsibility and retribution.

In chapter 9, I draw together my conclusions about free will and responsibility. I distinguish my position from other views generally favourable to free will, including views that free will is compatible with determinism, and views that suggest it is a special kind of causation, called *agent-causation*, that gives us free will; and I show how my position answers the 'luck swallows everything' argument against responsibility.

In chapter 10, I further advance my support for ideas of retribution by developing the idea that we can justly be praised or blamed (and even punished) for what we do, considering *whether there can be any truth in value judgments*—because if there is not, how could it be truly just to praise or blame anyone for anything? I also consider whether we have the capacity to make reasonable judgments about these matters. In chapter 11, I give reasons why I contend it is important to maintain ideas of responsibility and retribution in the criminal law.

Finally, in chapter 12, I consider what my approach to free will suggests about our universe and ourselves, concluding that our universe is not one of 'blind pitiless indif-

ference', as Richard Dawkins would have it; but that, on the other hand, there is no sound basis for accepting the literal truth of any religion, and indeed that fundamentalist religion of any kind has the potential for great evil.

As I proceed, I will highlight and number core assertions, and I will collect at the end of each chapter the core assertions of that chapter. This will I hope help to make clear the development and cumulative effect of the case made out in this book.

This book is a greatly expanded version of an argument previously set out in an article entitled 'Partly Free' published in *Times Literary Supplement* in July 2007. I believe it constitutes an original, coherent and plausible approach to a group of age-old unsolved philosophical problems, and the book is intended as a contribution to philosophy. However, I have tried to write it in such a way as to make it fully understandable by non-philosophers who are interested in questions such as whether or not we truly have free will, and whether (and if so to what extent) our responsibility for our conduct is affected by our very different genetic and environmental backgrounds.

I

Foundational Beliefs

CAN I BE CERTAIN THAT I EXIST?

Like René Descartes in the seventeenth century, I would like to start my enquiry with something I can be absolutely certain about. Descartes asked himself what he could not doubt, and answered to the effect that he could not doubt that he was asking this question, leading him to be certain about both his thinking and his existence: ‘I think therefore I am.’

As a young university student, I was quite impressed by the argument that the one thing I could be certain about was my own existence and experiences, and that everything else, including the existence of a world independent of me, and the existence of other persons who also had experiences somewhat like mine, was uncertain, supported at best by inconclusive inference. But I never embraced the view (called *solipsism*) that I and my experiences are all that there is; and I never believed that there was any real doubt about the existence of the ‘outside world’ and other people, because I thought the inference that these things existed was pretty strong. But I also thought this line of argument tended to support Descartes’s starting point.

Later on, I realised that Descartes’s argument could not establish very much about the nature of the ‘I’ that was said to be doing the thinking and that therefore existed. Certainly it could not establish that ‘I’ was an immaterial soul or self, distinct from my body, or something that could survive the death of my body. I still felt sure that for there to be thinking there had to be a thinker; but I saw no compelling reason why the thinker had to be non-material, or even had to persist beyond any particular occasion of thinking.

So I came to accept Descartes’s premise ‘I think’ as supporting no more than that ‘thinking is going on by something’, with the possibility that this something could have no more substance or longevity than the particular thinking in question.

Later again, I came to the view that even this formulation cannot be a satisfactory starting point as the most certain of my beliefs. This is because its truth is conditioned on the truth of certain assumptions, so that one cannot justifiably be more certain about this formulation than one is about those assumptions. There are two main reasons for this.

First, this formulation distinguishes *something doing the thinking* from the thinking itself, and thus presupposes that there is a thinker, albeit possibly an ephemeral one, which is not itself merely part of the thinking. But perhaps the apparent thinker is mere illusion, or a construction that is part of the thinking or at least not distinct from the thinking. Perhaps there is no *subject* of experience that can be distinguished in a satisfactory way from the *content* of the experience. This idea can be traced back to Hume's view of the self as a bundle of perceptions, and finds recent expression in the views of people like the philosopher Daniel Dennett¹ to the effect that our notion that there is a self doing the thinking is a device selected by evolution to help the brain make sense of its own processes. Now I disagree with these views, but they need to be addressed on a firmer basis than is appropriate at this stage of the enquiry.

Second, and more fundamentally, even the formulation 'thinking is going on by something' is an assertion the full significance of which can be appreciated only through a quite sophisticated understanding of the language in which the formulation is expressed; so that the truth of this assertion is conditioned on the truth of assumptions about the correct use of this language and the correct use and meaning of expressions within this language. How can I be certain that 'thinking is going on by something', unless I have sufficient understanding of this expression to appreciate precisely what it does assert and what it doesn't assert? This in turn requires an understanding of the relevant language and how it is used. It could be argued that all that is necessary is to understand just a few simple concepts; but while I will argue that some concepts are sufficiently simple not to raise questions concerning subtleties of language use, I contend that the assertion 'thinking is going on by something' does not have that kind of simplicity. It raises difficult questions about exactly what *thinking* is (as distinct from other kinds of mental activity), and exactly what it is for thinking to be *by something*. These questions can be explored only through careful use and understanding of the relevant language. And language presupposes a community of language users, and practices as to the use of language by that community. So a belief that thinking is going on by something cannot be justifiably held with any more assurance than a belief that there is a language in which words 'thinking' and 'by something' have a place and a meaning.

The same considerations apply, perhaps with greater force, to the assertion 'I think', because of difficult problems, mentioned above, associated with the meaning and significance of 'I'. That is, in addition to the problems raised by the question whether there is a thinker of any kind distinct from the content of the experience, there are problems raised by the question of exactly what is this 'I', what is *the subject of experience*.

¹ Dennett (2003), 245–55.

So far in this chapter, I've been explaining why something like Descartes's 'I think therefore I am' is *not* my most certain belief. What then *is* my most certain belief? I will set to one side beliefs that *there is something rather than nothing* and that *things happen*: the beliefs I am concerned with are beliefs with some *content* about what it is that exists and what it is that happens, and about how and why things exist and happen. The belief of this kind of which I am most certain is:

1.1. *Conscious experiences occur.*

Of course, I've expressed this belief in the English language, and this expression depends on there being such a language and on appropriate use of the words 'conscious', 'experiences' and 'occur'. But the idea expressed can be considered as quite a primitive idea not depending on subtleties of language use. Understanding of the idea does not presuppose sophisticated understanding of language use concerning *properties* of conscious experiences or *kinds* of conscious experiences or the existence of an experiencer, or concerning distinctions between types or modes of occurrences.

I'm not absolutely certain that non-human animals have conscious experiences, although (as will appear shortly) I firmly believe that at least mammals and birds do. Assuming that they do, I believe there is a sense in which an animal *knows* that an experience it has is occurring, for example that a pain it feels is occurring, even though of course it could not express this in language. It is in that general sense that I am certain above all else that conscious experiences occur.

When I feel a severe pain, I am certain that this pain is occurring. Whether it is correct to call it 'pain', or to say it is 'my' pain, or even 'my conscious experience', are matters depending in part on language and its correct use; and I cannot justifiably be more certain about these things than I am about the existence of language and of language users, and about the way language is used by them. But so long as I don't attempt to categorise types of conscious experience, and so long as I don't attempt to identify an experiencer, my belief in the occurrence of a pain that I feel can be considered as an acceptance of an idea that does not depend on subtleties of language use; and taken in that sense, my belief that conscious experiences occur is more certain than any other belief I have. And while I have selected pain as my first example of conscious experiences, the conscious experiences, of whose occurrence I am certain, extend to the whole range of conscious sensory, perceptual, cognitive and emotional occurrences that make up my ongoing conscious experiencing (although, I repeat, I do not include the categorisation of these experiences as part of this most certain belief).

Of course, the questions of what is consciousness, and what is its place in the world, are extremely difficult and controversial questions; but in this most basic belief I am not asserting anything about these questions. I am merely saying, about such conscious

experiences as I happen to have, that *these* occur, whatever they may be and however they may be correctly described.

This is the most certain of my beliefs, if only because there is a theoretical possibility that all my conscious experiencing could be some kind of a dream or illusion; although of course without language I could not say or believe anything useful about dreams or illusions, or contrast them with experiences that are not dreams or illusions but are experiences of reality. And in fact I do not believe for a moment that there is any real possibility that my conscious experiencing is no more than a dream or illusion.

Because this is the most certain of my beliefs, one important requirement that I look for in other beliefs is that they be consistent with and cohere reasonably well with this belief. It is conceivable that I could be caused to doubt or modify or even abandon this belief by some other consideration or combination of considerations; but this hasn't happened yet and I don't think it will happen. And although this belief might seem trite, vague and general, I think it is most important, because it is a belief about something that science has so far failed adequately to accommodate or explain. I will elaborate on this in chapters 7 and 8.

I also make the more general comment that, although I have gradations of confidence in my beliefs, and although in general terms I am starting in this book with near-certain beliefs and then progressing to less certain beliefs, I am not saying that my most certain beliefs are not revisable, or even that I would not revise beliefs of which I am presently more certain because of considerations associated with or derived from beliefs of which I am presently less certain. I am aiming for consistency and coherence; and although, where I find some lack of consistency or coherence, I generally look to modify beliefs of which I'm presently less certain rather than beliefs of which I'm presently more certain, this will not necessarily be the course I finally take. My ranking of beliefs in terms of certainty or assurance is itself a belief that is no more than provisional.

There is one further belief that I have about conscious experiences, in which I have the highest confidence and which does not depend on distinctions that require sophisticated understanding of language; and that is my belief that some conscious experiences *change*. Visual and auditory experiences, for example, are generally not static but are changing. Whether this is a matter just of changes occurring in the experience itself, or whether it indicates changes occurring in something that is being experienced, involves matters depending, in part at least, on categorisation and language; but the belief in change in the experiences is a primitive idea not depending on such niceties.

As with my first belief, it can be said that there are questions concerning *change* that are extremely difficult and require careful analysis through language: for example, the relationship between change and the passage of time. However, my belief that some conscious experiences change asserts no more than that, among those conscious experiences that I happen to have, some at least are not static or constant, but rather change.

I believe there is a sense in which an animal can know that an experience it has is changing, for example that sound it hears is not constant, even though of course it could not express this in language. It is in that general sense that I have high certainty that some conscious experiences change; and I would challenge anyone to come up with any reason for doubting this.

So I give as my second belief:

1.2. *Some conscious experiences change.*

This too might be considered trite, vague and general; but I think it too is important, particularly because it is not adequately accommodated or explained by a prominent scientific view about the universe and nature of time. I will return to this in chapter 7.

THE NEED FOR LANGUAGE

Now, once I attempt to go beyond the beliefs that conscious experiences occur and that some conscious experiences change, I find I cannot formulate and consider beliefs in any useful way without drawing distinctions that need to be identified with some precision in language. There is a sense in which I, like non-human mammals and pre-linguistic human beings, can have quite complex beliefs about the world that do not depend on their being expressed in language. For example, a lion may believe in the existence of its prey that it sees, and a baby may believe in the existence of a toy it is grasping, and these beliefs would not depend on language. But for me to achieve anything constructive in formulating my beliefs, and even in thinking about them, I have to use language and use it correctly; and this in turn depends on my having an understanding of how language is used by a community of language users.

My experience of life, my interactions with other people, my observations of their use of language, and my own use of language, give me strong reasons to believe in the existence of language and in the existence of communities of language users, of one of which I am a member. There is no good reason for me not to believe in these things, and this belief is necessary if I am going to make sound progress in expressing or even thinking about my beliefs. So I think it is reasonable to place next in my hierarchy of fundamental beliefs the following:

1.3. *There is language, and there are communities of language users, of one of which I am a member.*

This belief does not say much about the characteristics of members of the language community that includes me, other than that they have the ability to use the language

that I use. But it is a starting point for further beliefs in which I can be similarly confident.

The members of the language community that includes me are all human beings, one of many species of animal life on this planet. I confidently believe that I am a human being, one of a large number of human language users; although that is not to say that this is all that I am or that this describes exactly what is most important about what I am; or that this is precisely what I am referring to when I talk or think about myself as a subject of experience and action. An assertion that I am a human being reserves judgment on whether or not I may be an immortal soul, or at least a spiritual being of some kind, or a conscious subject or self that is not wholly dependent on my existence as a physical creature. So:

1.4. *My community of language users consists of human beings, a species of animal life, and I am a human being.*

EXPERIENCES AND THE EXTERNAL WORLD

Human beings have many characteristics in common, in terms of physical constitution, appearance, behaviour, and so on. The language that the members of my language community share can express and describe many features of ourselves and the world in which we live, *including conscious experiences of the general kind that are the object of my primary belief*. Because of our similarities, and because it is reasonable to believe that language developed and was used (and continues to develop and be used) to express and communicate things that language users could understand, it is reasonable to believe, with very high assurance, that these language users, and indeed all normal human beings, have conscious experiences of this general kind. But it is also reasonable to believe that the actual experiences that give me certainty that conscious experiences occur are experiences that *I* have; and that while other human beings have experiences that are of the same general kind as the experiences I have, their actual experiences are experiences that each of *them* has and I *do not* have.

1.5. *The actual experiences that give me certainty that conscious experiences occur are experiences that I have, and other human beings also have experiences of the same general kind.*

There also seem to be many features of the world in which we human beings live that are expressed and described in language, but are not themselves experiences of human beings, although they seem to be *objects* of experiences and potentially causative of experiences.

There is a vast philosophical literature concerning the place of conscious experiences in the world and their relationship with other aspects of the world; and there are many possible views.

One, called *idealism*, is to the effect that everything that exists is a kind of conscious experience or somehow reducible to conscious experience. Another, called *materialism* or *physicalism*, is to the effect that everything that exists is physical (of the nature of matter, energy and the like), and that conscious experiences are correctly considered as just one way of regarding or describing certain physical things or states or processes. Both of these views are species of *monism*, which holds that there is one basic mode of existence. Other views fall into the broad category of *dualism*, which holds that there are two basic modes of existence, that of conscious experience (or the *subjective*) and that of the physical world (or the *objective*), the latter of which can cause conscious experiences but is fundamentally distinct from conscious experiences.

There are two broad kinds of dualism: the dualism of *substance* associated with Descartes, drawing a distinction between immaterial souls or selves which have experiences, and material brains and bodies and other things that don't; and *dual aspect* theories, according to which at least some things or systems may have both a material or physical aspect and also a mental or experiential aspect. And within dual aspect theories, there are those that in effect say that there is a dual aspect to everything (*panpsychism*) and those that say that the capacity for experience is an aspect only of things like the central nervous systems of humans and animals, and thus has *emerged* only with the development of such systems.

And there are other theories that may be considered as denying all the above; in particular, theories saying that the physical and the experiential are manifestations of something else common to both (*neutral monism*), some of which assert that the distinction between the physical and the experiential is just one of many distinctions of similar importance that can be drawn concerning what exists.

Table 1.1 shows the general relationship between these views.

Some of the distinctions between some of these views are not sharp. In particular, there are versions of panpsychism that recognise gradations of the experiential aspect of things, from the full consciousness of human beings down to some minimal consciousness or potential for consciousness, and these versions can come very close to emergence theories. And the distinction between neutral monism and dual aspect theories is also not sharp, lying mainly in the degree of emphasis given to the distinction between the physical and the experiential.

I can't attempt to resolve the differences between these views at this stage of my enquiry or with the assurance required here. But I can make two preliminary observations about them, and express three further beliefs that I hold with high confidence.

Table 1.1 Views on the relationship between the experiential and the physical

<i>Monism</i>	<i>Dualism</i>
<i>Idealism</i> Everything is experiential	<i>Substance Dualism</i> Some things are experiential and other things are physical
<i>Physicalism</i> Everything is physical	<i>Dual Aspect Theories</i> At least some things have both a physical aspect and an experiential aspect:
<i>Neutral Monism</i> Everything is neutral as between the experiential and the physical	subdivided into <i>panpsychism</i> (everything has an experiential aspect as well as a physical aspect) and <i>emergence theories</i> (everything has a physical aspect but only some things have an experiential aspect)

First, and contrary to some versions of idealism, I'm confident there is an 'external world' that is not just part of the conscious experiences of individual human beings. Features of this external world are perceived similarly by different human beings, and in that sense are common to and accessible in common by different human beings, and so can be considered *objective* rather than being or depending on conscious experiences of any individual human being; and they are treated as such in our language. I am not here asserting that these features must be wholly independent of conscious experiences, merely that they are not just part of the subjective experiences of individual human beings but are objective in the sense of being accessible to and appearing much the same to different human beings.

There is an argument sometimes advanced in favour of idealism to the effect that we cannot conceive of anything except as being a possible object of conscious experience; but I don't think this in fact supports idealism. However, it does support something like Kant's view that we cannot know about how things truly are in themselves, but only about how they seem to us, with our particular perceptual and cognitive capacities; so that, for example, it may be said there is no such thing as colour in the external world, but only features that induce experiences of colour in normal human beings. I will say more about this later.

So I state my sixth and seventh fundamental beliefs:

1.6. *There is a world external to the conscious experiences of individual human beings, features of which are accessible in common by human beings and describable in language.*

1.7. *Our language may describe at least some of these features by reference to how these features seem to normal human beings, and this may not adequately describe how they are in themselves, independently of any experience of them by us.*

Second, I am certain that human beings who are awake and functioning normally have conscious experiences like mine, and I believe with high confidence that other mammals and also birds have the capacity for conscious experiences somewhat like mine, because of their similarities to human beings in physical constitution, origins and behaviour. I believe with decreasing confidence that much the same is true of reptiles, amphibians, fish and invertebrates. I think it unlikely that either plants or more primitive forms of life have that capacity. I believe with very high confidence that anything that is not alive, such as a rock, does not have conscious experiences in the same sort of way as humans and at least some other animals have experiences. Thus with high confidence I believe:

1.8. *Some entities or systems (including humans and some non-human animals) have conscious experiences more or less like mine, and some (including rocks) do not.*

This is not an outright rejection of panpsychism. I am not asserting that there could not be some primitive experiential aspect to non-living things, merely that they are not conscious in the sense of having experiences that are in any way similar to the experiences that human beings have, and that (I believe) some non-human animals also have.

There are some scientists and philosophers who contend that some non-living systems, such as computers, are or may be conscious in this sense. In chapters 4 to 6 I will argue for views indicating that computers, in their present form, could not be conscious; and that the exponential increases in computing power that are expected to occur in coming years will not make them conscious. So in that crucial respect, I suggest, computers are more like rocks than like human brains.

FOUNDATIONAL BELIEFS

1.1. *Conscious experiences occur.*

1.2. *Some conscious experiences change.*

1.3. *There is language, and there are communities of language users, of one of which I am a member.*

1.4. *My community of language users consists of human beings, a species of animal life, and I am a human being.*

1.5. *The actual experiences that give me certainty that conscious experiences occur are experiences that I have, and other human beings also have experiences of the same general kind.*

- 1.6. *There is a world external to the conscious experiences of individual human beings, features of which are accessible in common by human beings and describable in language.*
- 1.7. *Our language may describe at least some of these features by reference to how these features seem to normal human beings, and this may not adequately describe how they are in themselves, independently of any experience of them by us.*
- 1.8. *Some entities or systems (including humans and some non-human animals) have conscious experiences more or less like mine, and some (including rocks) do not.*