

Preface

About this Book: Why the world in my mind?

There is no shortage of books on consciousness, so why write another one? Over the last five years since I finished my last book¹ something has become blindingly clear to me. Consciousness is many things: at least, as we shall see, I have identified five that seem crucial. A scan of the many of the excellent books that claim to explain consciousness does not reveal systematic attempts at breaking consciousness down into simpler elements to make the concept more accessible. This is like trying to explain how a grandfather clock works without referring to the pendulum and weights. I first published this five-step idea as a journal article² based on a contribution to a small, fascinating meeting at the Cold Spring Harbour Laboratories in 2001 (http://www.swartzneuro.org/banbury_e.asp) organised by Christof Koch, David Chalmers and Rod Goodman. The question was 'Can a Machine Be Conscious?' and there was a surprising degree of agreement that one could – a notion that has driven my research since the late 1980s. But looking at the arguments put forward by myself and others made me realise that the level of technical detail we use to make our respective cases is dire. This prevents communication not only among the experts but certainly between the experts and anyone out there who

[1] Aleksander, I. (2000), *How to Build a Mind: Dreams and Diaries* (London: Weidenfeld and Nicolson; paperback ed. Phoenix 2001).

[2] Aleksander, I. & Dunmall, B. (2003), 'Axioms and tests for the presence of minimal consciousness in agents', *Journal of Consciousness Studies*, **10** (4–5), 7–19.

would like to know more about consciousness and join in the debate.

Koch talks about the neural correlates of consciousness and coalescence, Chalmers of something called the logical non-supervenience of consciousness on the physical brain and Goodman with others argues that model-referenced control systems are the answer. I talk of five formally stated axioms. This, therefore, is why I have written this book. All this specialist mumbo jumbo does map into normal language. And once fixed in normal language it may be used to answer questions that many of us have about the nature, origin and use of consciousness.

I start with five *axioms* — there we go, why use a word such as ‘axiom’? What is it? An axiom is a plausible idea, rather than a proven truth, on which one can build sensible explanations of things. So being conscious for me breaks down into five basic ideas that raise important questions to which I try to provide plausible answers. Here are the five axioms or steps of this book and the questions that they prompt. The book is about the answers.

- The major part of being conscious is my sensation of **being an entity in an out-there world**. How does this happen?
- Another part of being conscious is that **I am an entity in time**: I have a remembered past and I can take a few guesses about the future. What mechanisms have this property?
- It is remarkable that we have similar brain mechanisms but they lead to different personalities and different ways of being conscious. Something called ‘**attention**’ is at work here: what is it and how does it work?
- My mind seems constantly engaged in sifting choices about **what to do next**. How does it do this to my best advantage?
- I seem to be influenced and guided by things I call **emotions**. What are they?

Since the mid 1990s, I also have written about the way in which machines could be conscious.³ I came to the conclusion that, more than a non-human animal, a conscious machine could come close to discussing its own consciousness. This does not mean that machines will compete with us in late night conversations of whether this or that philosopher is right. It does mean, however, that, having got close to the mechanisms of consciousness through the design of machines, helps to answer the axiomatic questions I raised earlier. So the ideas in this book are driven by two strong impulses. The first is that I need to be outrageously introspective. I shall talk of those strange things that I feel personally as indicators of what needs to be explained. The second is that explanations should be cast in terms of simple principles that anyone can understand. I don't want to give the impression that I am the first to have given thought to this type of deliberation. Indeed, some of the book discusses the fascinating and diverse thoughts of others, but my personal experience and what mechanisms might be involved are the twin schemes through which the five steps to being conscious are taken.

The first chapter (**Capturing the Butterfly of Thought**), owes its title from an early task we set our models of conscious mechanisms: to imagine a butterfly. This got quite a lot of undue publicity along the lines of being a 'break-through' in building the world's first conscious machine. Nothing could have been further from the truth: the machine called Magnus *was not* conscious but was the first dedicated piece of software that enabled us to study hypotheses about the way that consciousness might emerge from brain structure.

This chapter asks what is to be gained by building machines and how does this help those who do not generally build things? First, building things requires a clear definition of what will be built. Words like *thought*, *consciousness* and *mind* are given an operational character. Second, it establishes that 'being conscious' is about *mechanisms* that make us conscious which is a saner way to go than wondering about '*having consciousness*'

[3] Aleksander, I. (1996), *Impossible Minds: My Neurons, My Consciousness* (London: Imperial College Press).

which promulgates the mistaken idea that consciousness is a property like having a leg or a cold. The work and ideas of others who build things are also discussed.

How do we decide that one thing is conscious and another is not? In the second chapter (**The Five Tests for Being Conscious**), I side with those who say that it is virtually impossible to tell if an organism is conscious just by observing the thing behave. But digging deep into ourselves we know what major sensations would seriously affect our being conscious were they to be missing. This is where the five axioms come from and this is the chapter in which they get a thorough airing. These then become the targets for identifying which parts of the brain are involved in the axiomatic phenomena and what kind of computational machinery could possibly do the same thing. Luckily it is not necessary to have a PhD in computer design to understand the arguments that are involved in looking at being conscious from the point of view of 'how could a machine do it?'. This is precisely where the notion of 'capturing the butterfly of thought' comes in. I think of a butterfly but wonder what it is in the activity my brain cells that makes it into a butterfly. If the cells of a machine can do the same thing, have a sense of self, attend to important things in its world, plan and evaluate its plans, we will be well on the way to being much clearer about what it is to be conscious.

An exceedingly strange thing that my brain does is to deprive me of the power of thought about once a day, usually at night. Chapter 3 (**Sleep, Dreams and the Unconscious**) is about the relationship between what we call being conscious, its loss when sleeping, its peculiar reappearance when dreaming and its disappearance while being anaesthetised. We also hear stories about something called 'the unconscious'. It may seem odd that having hardly got into what it is to be conscious I slip into discussing what it might be to be unconscious. Perhaps this is not so strange given that understanding the difference between the two makes *being* conscious stand out from *not being* conscious. What is sleep for? What is the difference between what goes on when I am awake, when I sleep and when I dream? After all during the latter my brain is fully alive and working. Here we begin

to flex the muscles of the axioms of chapter 2 to give us some insights into these areas of our existence in which the 'self' appears to have taken a holiday. We look at what is known from brain-wave measurements and then take a look at a little simulation that throws light on what sleep experts call the 'sleep cycle'. Axioms are also helpful in clarifying what Freud might have meant by 'the unconscious'.

In Chapter 4 (**The Octopus with a Stomach Ache**) I wonder whether non-human animals are conscious. Experiments on the octopus show a remarkable ability to plan and assess long sequences of events. This is just one example of what makes the question of animal consciousness one of the most passionate debates in the science of consciousness. I look at the struggles of scientists who are designing experiments to discover glimmers of consciousness in animals. They surely are up against it. It is difficult enough to infer consciousness in people one knows well: the problem is immensely difficult when it comes to animals. But while the experiments mostly do not provide fool-proof evidence of thinking as we know it, they virtually never cut some form of deliberative mental state out. In terms of the axioms of this book, theory has it that it is likely that animals are conscious — not of exactly the same things as we are, but conscious of what is important to *them* nonetheless. The axiomatic argument centers on the similarity of mechanisms between man and beast and the likelihood of such mechanisms doing similar things for both.

Can it happen that a gorilla can walk about in a crowd without being noticed? Yes it can as has been shown by a simple experiment that involves what psychologists call 'attention'. Those taking part in the experiment are asked to count the number of bounces of a ball in the film of a ball game played among a few young people. Only less than 20% notice a man in a gorilla suit when he crosses the playing field in a rather obvious way. Chapter 5 (**The Disappearing Gorilla**) examines this essential but peculiar behaviour of the brain. This kind of experiment leads some to believe that our feeling of being an entity in an out-there world is nothing but an illusion. I side with those who argue that the illusion idea is mistaken. Attention, which some call 'the

gateway to consciousness' is a very useful device for selecting what gets into our consciousness out of a flood of sensory information. The process of selection is complex and sophisticated, forming the meat of this chapter. It is also what causes differences between us. No two people attend to the same things. This is why *some* people see the gorilla. Part of our personalities is determined by the way we attend to the world around us. But without attention we would drown in sensory information. So there can be situations in which the attention mechanism leaves things out. The axioms suggest that this is an exception and not a rule. Our vision is no illusion it is just not always perfect. This discussion also allows me to visit a theory recently put forward by Alva Noë of the University of California at Berkeley and Kevin O'Regan of the René Descartes University in Paris. They argue that the body, including the eyes, reacts to the world in an unconscious, automatic fashion. They see attention as the process of breaking into this process and thus becoming conscious. This would offend our axiom that deals with imagination (axiom2) as it is based on the idea that all the memory an organism needs is out there in the world. But then, how could we remember the past? I argue that a reconciliation of various points of view will make sense of that vital part of our consciousness: sensing the external world.

Illusions are fashionable in the modern philosophy of consciousness. In Chapter 6 (**Knowing What We Want**) I look at a philosophy that suggests that we kid ourselves when we think that we first want something and then go out to get it. This is based on an experiment carried out by Benjamin Libet which showed that participants asked to lift a finger had brain activity associated with this event *before* they became conscious of what they wanted. Who is pulling the strings? This impinges on issues of free will which have been central to our cultures for as long as one can say cultures existed. I take a look at some of the history of free will and then use some of the ideas of Chapter 2 to show first, that the illusion arguments leave out the brain activity due to the emotion which is involved when we make choices. When this is included, the illusion idea becomes less credible and reasons for our feeling of control over what we want are reinstated.

This book advocates strongly that there is a tight link between some neural activity in the brain and personal sensation. This appears to contradict the philosophy of David Chalmers, an influential contemporary consciousness philosopher who created the notion of the ‘hard’ problem of explaining sensation. That is, whatever you understand at a physical level does not bridge over to what you sense. In Chapter 7 (**Chalmers’ Two Minds**) I take a close look at the hard problem and the explanatory gap between sensation and brain activity. Surprisingly, axiomatic arguments actually agree with the way Chalmers thinks that the gap might be bridged: through the use of informational and computational arguments.

Finally, in Chapter 8 (**Unfinished Business**), I suggest that this book is merely the beginning of what could be a level-headed way of removing a fear of mechanistic discussions of consciousness without losing the awe and respect that it commands. Here I briefly dip into some further important issues that are affected by the axiomatic arguments of this book. The idea of using computation to understand consciousness is under sustained attack. One of the most elegant assaults comes from geriatrician Ray Tallis and I look at the way that the axioms provide a defence.

Notable for its absence in the first seven chapters is the mention of natural language. Despite the very human (as opposed to ‘animal’) nature of this, I look at the way axiomatic mechanisms in humans might have this added facility over other animals. The ideas of Ludwig Wittgenstein and Noam Chomsky come into view while the axiomatic approach suggests structures that might resolve some questions that arise from these existing philosophies: consciousness of abstract ideas and whether language is inherited or acquired.

Probably the most important application of what is being said in this book is the provision of operational models for medicine as it relates to mental illness. Most mental illnesses are distortions of consciousness. For example, the memory problems of an Alzheimer’s sufferer may be expressed as failures of imaginal machinery while something dire has gone wrong with the axiom 1 depictive machinery of a schizophrenic. I

describe some work done with Helen Morton with Parkinson's sufferers. on visual planning deficits. Why should a lack of dopamine which normally affects movement interfere with working out how to move a few coloured spheres around? The axioms suggest an explanation.

Does the prospect of a mechanistic explanation of consciousness offend the religious beliefs of some? In this final part of the book I suggest that in the long term it should not. It is merely another example of the way in which theology needs to be the philosophy for things that we do not fully understand, but must then give up this hold as science provides rational explanations. So as a moral of the story of this book, I suggest that the mechanistic steps we take, help us to appreciate the sophistication and subtlety with which the brain gives rise to our rich, exciting and highly effective inner life.

Appendix

This is included for those who want a brief technical introduction to the details of the kind of neural machinery which is used to create simulations of conscious systems. This culminates in a 'kernel' architecture that is used in various parts of the book.

Thanks

The ideas that fill this book do not occur as a result of monastic isolation. First I am grateful for the sustained discussions and email exchanges with my colleague Barry Dunmall. Despite being right in the thick of the laboratory work on conscious machines, he remained challengingly skeptical which certainly encouraged me to be clear about my assessments, assertions and axioms. I wish to thank my current doctoral students whose work in machine consciousness leads to progress: Kirk De Souza, Peter Liniker, Sunil Rao, Mercedes Lahnstein and Rabinder Lee. The title of this book owes its inspiration to Max Velmans who recently gave his professorial inaugural lecture entitled 'Is the World in the Brain or The Brain in the World?' His clarity of thought and thorough assessment of the philosophy of

consciousness makes me feel happy to have been able to paraphrase the title of his lecture.

Thanks are due to the Leverhulme Trust whose help has made it possible to find time to write this book and to Imperial College for continuing to support my academic existence as an emeritus professor who finally can enjoy his research without worrying about external assessments and the extraction of funds from funding agencies.

Finally I thank Helen for continuing to show me how pleasurable life is away from the computer screen, Joe and Melissa for being a smashing couple and bringing the delightful Luke and Tai into vivaciously conscious life, and Sam and Claudine for being another smashing couple who make the best *pañ de queso* in the world.