So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.

There was nothing so very remarkable in that; nor did Alice think it so very much out of the way to hear the Rabbit say to itself, ‘Oh dear! Oh dear! I shall be late!’ (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually took a watch out of its waistcoat-pocket, and looked at it, and then hurried on, Alice started to her feet, for it flashed across her mind that she had never before seen a rabbit with either a waistcoat-pocket, or a watch to take out of it, and burning with curiosity, she ran across the field after it, and fortunately was just in time to see it pop down a large rabbit-hole under the hedge.

In another moment down went Alice after it, never once considering how in the world she was to get out again.

(Lewis Carroll 1865, Chapter 1: ‘Down the Rabbit-Hole’)

*I am me and you are you*: a banal statement that is hardly worth writing a book about. Yet, in the paraphrased form of *I have me-ness and you have you-ness*, it suddenly begins to raise interesting questions. What is the nature of the *me* that *I* can appreciate as being me? How does the *me* relate to the *I* that is recognising the *me*? Are me-ness and you-ness the same thing looked at from different angles, or is there an important difference between the two? How does the *you* relate to the *I* that is recognising the *you*? And what is the nature of the *you* that *I* identify as being you; is it the same as the *me* you identify as being you? All of the questions raised here are from the first-person perspective because, in
the end, it is the only perspective each of us has; but is it the only perspective we use? And, if not, how do I incorporate the perspectives of others into my view of the Universe?

In fact, having a self of which I am aware is perhaps one of the most astounding and unexpected outcomes of being human. Current scientific evidence seems to indicate that it is unusual in nature for an organism to be able to recognise itself (although the number of species able to pass the mirror test of self-recognition is growing constantly – see Chapter 2); and we have no evidence that any individual of any species – apart from humans – is able to imagine how others might see them. Our personal relationship with our selves may even be the 'holy grail' that has been sought for centuries: the thing that makes us different from other animals. If such a difference really exists, selfhood would seem to be a good candidate.

We know our species is different from others in important ways: every species is a particular outcome of a series of challenges to its individuals, such that individuals with strategies to meet the challenges do better than those without such strategies. After enough time, the species consists of only those individuals with useful strategies; and it is those useful strategies that define the nature of the species. Charles Darwin (1859 [2001]) formalised this understanding in his theory of descent with modification by means of natural selection, and Herbert Spencer (1864) summarised it, somewhat controversially, with the phrase *survival of the fittest*. You can identify the challenges that a species has met by looking at what its members are good at: for instance, the challenge of surviving predation has, in different species, resulted in climbing or running or hiding or fighting skills. Monkeys climb, antelope run, stick insects hide (in plain sight) and porcupines fight – or, at least, they are equipped with an effective active defence mechanism with which to discourage predators. So what are humans good at? And how does having self-awareness help us to be good at it?

First, humans are clearly very good at cooperating; perhaps not as effectively as the eusocial insects (ants, wasps, bees and termites) but certainly more effectively than any other primate. Second, we have language – a communication system that may itself be unique in nature, and which seems to be both an outcome of cooperation and a cause of even greater cooperation. Third, compared to other primates, we are abnormally willing to work together in joint enterprises that require specialisation and role-taking. Fourth, and most mysterious of all, we seem to be happy to subordinate our own needs to those of others, often to the point of self-sacrifice: we are willing to die to keep others living. Eusocial
insects also sacrifice themselves; but that is because they only get their genes into the future by keeping their reproductive parents and siblings alive. When a non-reproductive eusocial insect sacrifices itself, it does not disadvantage itself reproductively, and often it advantages itself by protecting its fertile relatives: the self-sacrifice of the non-reproductive individual does not contradict their evolutionary self-interest. In contrast, when humans sacrifice themselves, they forego future reproductive opportunities. This could be viewed in some circumstances as somehow advantaging their offspring, but humans often self-sacrifice before they have even had the chance to reproduce – which looks, in evolutionary terms, completely nonsensical. Could the capacity to imagine ourselves as having a self somehow be behind this willingness to self-sacrifice? If this is the case, it only leaves us with a different evolutionary conundrum: if having a self is implicated in such an evolutionarily unfit activity as self-sacrifice, how has the peculiarity of human self-awareness survived the inevitable evolutionary extinction that self-sacrifice entails?

About this book

This book looks at human selfness as the outcome of evolutionary selection: what, in our evolutionary history, made having a self a fit strategy? Humans have selves, and we consider having a self to be mostly a Good Thing (as Sellar and Yeatman 1930, would say); but having those selves can often make us unselfish and willing to subordinate our self-interest to that of others – which, in evolutionary terms, is a Bad Thing. We cannot convert our Darwinian self-interest into self-disinterest unless we become aware that we have a self that has an interest in its own survival; but how were we able to maintain our awareness of our selfness when doing so made us less likely to survive than our selfish neighbours? There is clearly an evolutionary tale to be told here.

This book also looks at the role self plays in language. We can self-reference – a capacity not unknown outside of our species, but rare. We can also model ourselves into a range of circumstances – not just in the factual present, but into the future, the past, the might-have-been and the maybe-will-be. We can even model ourselves in the never-has-been and the never-will-be – once again, a capacity that seems evolutionarily pointless. We can also, through language, share our knowledge of our self with others; and we seem more than happy to do so, even though it further reduces our relative fitness by giving away information that others can use against us. Language could be, like many other communication
systems, either non-volitional or strictly about external facts; but it is actually, compared to the communication systems of other species, highly volitional and often about inner cognition, especially what we know about our own and others’ selves. Strangest of all, the fact that we are able to choose to share own-self and other-self knowledge seems to be all the reason we need to do so.

Finally, the book will look at the models of self we hold in our heads. What are we modelling when we model a self? Is there a difference between our models of our own self and our models of other selves? And how did this capacity to model our selves survive and thrive in humans, when it seems not to play a significant or useful role in the lives of other animals? Does the capacity to model our selves mean that the only self of which we can be aware is actually a model, or is there something more basic, substantial and actual on which we build our models?

This book proposes a new hypothesis about selfhood, which I call the Seven-Selves Modelling Hypothesis (SSMH). The argument for the hypothesis is given across eight chapters, the first of which looks at some of the existing theories of selfhood. The chapter discusses religious viewpoints and the approaches of various key philosophers, what leading psychologists think, what neurologists have found about selfhood as a cognitive phenomenon and what anthropologists have observed about selfhood in human individuals and groups. By sampling the wide range of ideas about selfhood available in the literature, the chapter shows that the question ‘What is a self?’ still has no single answer from any single discipline. Perhaps a new, cross-disciplinary approach will prove more productive.

Chapter 2 approaches the question of selfhood from a different direction: where did it come from? It tells a story about the evolutionary development of selfhood from single-celled animals to modern humans, showing that it can be seen as the outcome of a series of developments in sensing and cognition about self and other individuals. Conscious awareness is a key event in the evolutionary process leading to selfhood, creating new ways for individuals to interact and new tools, such as Theory of Mind (ToM) and language, to facilitate the interaction. The chapter looks at the capacities for self-awareness in other animals and considers how human self-awareness may be different.

Chapter 3 concentrates on more recent evolutionary events to show how modern Homo sapiens evolved to be able to model a personal self. It shows that a necessary precursor is the capacity to make models of other individuals and the relationships between them – something that
requires a special rule-driven system. Here I use Derek Bickerton’s (2002) term, social calculus, to label this system. The chapter explores how this social calculus could have arisen, how it became shareable through language and how the grammatical complexity of language corresponds to the systemic complexity of social calculus. It also considers some of the particular features of language that closely reflect those of social calculus, and how the sharing of social modelling requires a communication system that is mainly used not for the sharing of truths and facts but for the sharing of opinions – a mode of communication for which language is especially suited.

Chapter 4 looks at how humans develop self-awareness in childhood: we are not born with it, but rather it is something that develops progressively through our childhood. The chapter considers developmental and social features that mould human children in a species-specific way, in particular our extended childhood and the extended caring network that supports it. A range of current theories of childhood are examined in relation to cognitive capacities such as delayed gratification, deception and self-expression.

In Chapter 5, the genetic and cognitive origins of social calculus are examined in greater detail. If self-awareness relies on the sharing of social calculus, then the origins of social calculus are a significant aspect of selfhood. The chapter looks for signs of social arithmetic and social calculus among a range of non-primate social species – parrots, corvids, naked mole rats, meerkats and bottlenose dolphins – before examining the development of social calculus in our own evolutionary clade. The path to the sharing of social calculus is traced, from the social arithmetic in the Machiavellian intelligence of chimpanzees, through various forms of altruistic behaviour (kin selection, reciprocal altruism, indirect reciprocity, costly signalling, altruistic punishment, vigilant sharing and reverse dominance) and on to the outcomes of shared social calculus, such as our capacity for self-sacrifice.

Chapter 6 looks at the role of self in language, and the role of language in sharing modelled selves. The linguistic and socialising roles of pronouns and pronominalisation are explored in relation to a selection of the world’s languages, and the use of names as personal labels is discussed as a route into pronominalisation. The origins of the three linguistic persons, they, you and me, are also considered – as markers of selfness, possession and reflexivity. The extended self, indicated by possessive pronominalisation, and the recursive self, indicated by reflexivity, are also analysed. Finally, the chapter considers how selfhood and language synthesise to increase communicative complexity.
Chapter 7 examines the importance of metaphor within language, looking at five key conceptual metaphors of selfhood and self-modelling. The first of these, THE MODEL IS THE ACTUAL, shows that we treat our social modelling as if it were a calculus of actual relationships between members of our group, even though it is just a representation of a set of opinions. The second metaphor, THE GROUP IS AN ENTITY, lets us treat a group as if it had the same motives and purpose as an individual; and the third, SELF IS OTHER, treats both my self and your self as third-person constructs, slightly more privileged than *they*, but essentially the same as *they*. The fourth metaphor, I AM ME, equates the objective self (*me*) with the active and interactive self (*I*): acting and being acted-on do not create different self-models, they are functions of a single self-model. The fifth metaphor is a little different from the others: ONE AMONG EQUALS reduces the status of my self as represented to myself, making it no more important than other selves as represented to myself; we self-police our own humility and obedience to the group.

Chapter 8 sets out the seven selves of the SSMH, showing how they work together to create our sense of selfhood. First, there is the unknowable Actual self, the genetic but subliminal recognition of the importance of the self to the self. Next, there is the Social self, the self others believe me to be; this is a self of which I am consciously aware, and it is generated from the social models of me that others have shared with me. The third self is the self-model, the self I believe me to be; it is my own conscious model of me generated from other people’s models of me. Fourth is the Episodic self, my self as modelled in my individual memories; and fifth is the Narrative self, my self as a continuous entity through time, the story that links the individual Episodic selves. The Cultural self is the sixth self, the self I *should* be; it is the model of the perfect citizen offered by others in the group, the best me I can be. This leaves the final self, the Projected self, the self I want others to believe me to be – which may only vaguely resemble my own self-model. The chapter also explores how these selves operate together to define our selves to ourselves and to our group.

In this book, we follow the white rabbit of selfness down the rabbit-hole of self. We are entering a strange universe where nothing is quite as it seems (and even *nothing* is not quite as it seems); and, once we have entered, we will not leave unchanged. Sometimes our self will seem very large and complicated, at other times it will seem to shrink and may even disappear. One moment we will be running as fast as we can just to stand still, the next we will find things changing around us without apparent logic or reason. I hope that you will find this journey informative or, at least, enjoyable – it is, after all, about you. And I hope you will come to
realise, by the end of this book, that there is considerably more down the rabbit-hole than white rabbits:

If any one of them can explain it,’ said Alice, (she had grown so large in the last few minutes that she wasn’t a bit afraid of interrupting him,) ‘I’ll give him sixpence. I don’t believe there’s an atom of meaning in it.

(Lewis Carroll 1865, Chapter 12: ‘Alice’s Evidence’)