Origins of Self

Edwardes, Martin P. J.

Published by University College London

Edwardes, Martin P. J.
University College London, 2019.
Project MUSE. muse.jhu.edu/book/81330.

For additional information about this book
https://muse.jhu.edu/book/81330

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=2771966
'... And I haven’t sent the two Messengers, either. They’re both gone
to the town. Just look along the road, and tell me if you can see
either of them.'

'I see nobody on the road,' said Alice.

'I only wish I had such eyes,' the King remarked in a fretful
tone. ‘To be able to see Nobody! And at that distance, too! Why, it’s
as much as I can do to see real people, by this light!’

(Lewis Carroll 1872, Chapter 7: ‘The Lion and the Unicorn’)

What is happening in this quote? Is the White King just playing with
words, or does this dialogue represent an important aspect of self-
awareness – at least, in language? The king asks Alice if she can see a
specific thing (either of the two messengers); Alice’s reply, though, is
ambiguous: is she saying that she cannot see the specific thing (either
messenger), or that she can see a generic thing that doesn’t actually exist
(‘nobody’)? Either way, she has not answered the king’s question directly
and clearly enough, prompting his counter-response about the capacity
to see things that are not there.

And yet we tend to consider Alice’s contribution to the conversation
as normal, and the king’s response abnormal. Practically and pragmatic-
ally, there is no difference in meaning between ‘I don’t see anybody’ and
‘I see nobody’. In fact, it is only when the presence of absence and the
absence of presence are different states that the king’s problem appears: ‘I
don’t see any me’ and ‘I see no me’ can be the difference between having
a self you cannot identify in the current context and not having an iden-
tifiable self at all. It is this difference that lies at the heart of the modern
debate about selfhood.

So which approach is right? Is it difficult, or impossible, to see my
self myself, or is there no self to see? It depends, as we saw in Chapter 1,
on which theory of self you start with. Is a self a collection of predictable
responses that produce selfhood as an identifiably individual thing? Or
is self a process of responding predictably that allows others to see you
as an identifiably individual thing? And, if the second is true, does this
mean that my self is an actual thing, a thing we can agree is real, or a
convenient illusion?

As Chapter 2 made clear, having a self you can be aware of is
problematic in evolutionary terms: it seems to reduce the fitness of the
organism by allowing it to adopt some very un-Darwinian strategies. Yet
not having a self does not fit well with the cultures we live in: even the
most self-denying human cultures believe that the self can be called to
account and punished. We may not actually have a self we can be aware
of, but we still have to act as if we do! So, between these contradictions,
is there a middle road that can explain both the problematic unfitness
of being self-aware and the social insistence on continuous selves? The
modelled self may give us just such a middle way.

How to make models of others

To discover how we make models of ourselves, we must start with our cap-
acity to make models of others. This is a capacity that lends itself well to
an evolutionary explanation, and which seems to be widespread among
our fellow primates and beyond. The individual who can anticipate the
actions of their conspecifics, their predators and their prey will have
an advantage over the individual who cannot, and they should there-
fore be better at surviving and thriving (Seyfarth and Cheney 2013). Of
course, an advantage often comes with some countervailing disadvan-
tage; and, in this case, the capacity to make social models comes with
the significant energy costs that the necessary enhancement of cognition
requires. However, as long as the advantages of a capacity outweigh the
disadvantages, that capacity will be selected for in evolutionary terms;
and, as the capacity to anticipate the actions of others has become com-
monplace in many species, it clearly must be advantageous.

The advantages of social modelling in the predator–prey dyad are
different from those between conspecifics. Between predator and prey
there is an evolutionary war of strategy and counter-strategy to get an
individual’s genes into the future. Usually, a single strategy and a single
counter-strategy define the relationship, and both are emphasised and
enhanced by natural selection over the generations. While it does not
involve social modelling, an example of the strategy and counter-strategy
process is the golden poison arrow frog, *Phyllobates terribilis*, which is in an arms race with its only predator, the *Liophis epinephelus* snake. The frog has a skin neurotoxin that is so powerful that it can kill a human just through contact, which seems like a massive overkill (literally). However, over time, the snake (whose diet comprises largely golden poison arrow frogs) has developed a tolerance of the toxin. So the frogs with higher toxicity survive better than the less toxic, while the snakes better able to deal with the poison survive better than the less capable. This is Matt Ridley’s (1993) Red Queen problem in action: in evolutionary terms, both the frog and the snake are running as fast as they can just to stay still.

The social modelling strategy, in contrast, gives humans an advantage that is difficult for both our predators and our prey to counter: the advantage of group cooperation. Because social modelling allows humans to anticipate each other’s actions, both offensive action against prey and defensive action against predators can be coordinated. However, coordinated action against predators and prey is not an unusual capacity in nature, and is not limited to socially clever species: babbler finches coordinate mobbing effectively against predators (Zahavi and Zahavi 1997, 5–6) and wolves coordinate the killing of a range of prey (Baan et al. 2014). Even eusocial insects like ants and bees show a high level of group cooperation in defending against threats to their nests (Whitehouse and Jaffe 1996). It does not take a high level of cognitive sophistication to produce the social cohesion needed for tactical defensive or offensive cooperation.

Between conspecifics, though, the attack/defence arms race needs a more sophisticated competitive process, because the bidding war between a single strategy and a single counter-strategy is less efficient. For instance, the poison arrow frog has to have a tolerance to its own toxin, so its conspecifics will also have this tolerance: the toxin is useless in a competition with another poison arrow frog. For competition between conspecifics, it is better to have a range of strategies and a capacity to switch to a new strategy if the first one doesn’t work. In this case, communicative strategies become more significant. For instance, one of these new strategies will be submission, a signal indicating that the losing party in a contest is giving up. As a general rule in conflicts between conspecifics, it is to neither party’s advantage to continue a fight beyond victory; so it is in the interests of the victor to recognise a surrender – and, therefore, in the interests of the vanquished to be able to signal surrender.

This is an example of what Thom Scott-Phillips (2015) describes as a code model of communication: there is a signal that is advantageous
for the sender to make and for the receiver to acknowledge, so the sender and receiver evolve toward each other in terms of signal and response. If the surrender is to work, though, the sender and receiver must both be able to model the likely reactions of the other party. It is important to the sender that this particular receiver will recognise and act on the signal, and it is important to the receiver that this particular sender is honest in their surrender. The reliability of the signal is founded on the intentions of the sender and receiver, so it becomes useful for both parties to be able to model each other’s motivation, and to have volitional control over when to signal and how to react to a signal. This is what Scott-Phillips describes as the ostensive–inferential model of communication, which ‘involves the provision and interpretation of evidence for the meaning that the speaker intends to convey’ (2010, 95). Or, to put it another way, human communication is an immediate negotiation toward meaning between two people, not a slow, evolutionary negotiation of signal and response.

If a signal is volitional, however, it can be subverted. Richard Byrne (1995, 125–6) tells of an adolescent baboon, Melton, who had played with an infant too roughly. The infant, naturally, screamed for its mother; and she, with several other adults, began to chase the hapless adolescent. However, instead of trying to outrun the adults, Melton stood up and began to scan the horizon – a signal to other baboons that a predator has been spotted in the distance. The adults ceased chasing and also began scanning, looking for the threat that had cued Melton’s behaviour. Melton had distracted the adults from their punishment detail with a more serious – although non-existent – threat. His behaviour could be read as a deliberate subversion of the predator signal, and a deliberate deception of the adults; or it could be read as a transferred fear reaction, causing him to adopt a pose that had helped reduce his fear in the past. Either way, the predator warning had been subverted and the adults deceived.

Melton shows us that, in primates at least, the link between the mental state and the external signal produced by that mental state is more complicated than just a simple stimulus–response system: the external signal is made only if the signaller decides (deliberately or automatically) to make it. The external stimulus has to become an internal perception, and a judgement has to be made about its relevance. This judgement may be faulty, leading to a signal being made inappropriately, even if it is not under deliberate control; but, if the sender has deliberate control over producing the signal then an extra level of deception, intentional deception, becomes possible.
This intentional deception relies on the signaller understanding, on some level, the effect the signal will have on the recipient. The signaller has to be able to model the recipient as a motivated object – not necessarily a being with its own agenda, but an object that is likely to respond to a signal in a reliable way. The signaller must also understand the advantage that the deception will give them over the recipient, and how to capitalise on that advantage. It requires the signaller to have a versatile understanding of the likely responses to the signal, and of the various ways to exploit those responses.

The ability to model others as motivated objects is not limited to humans, or even primates; it seems to be an ancient faculty, occurring in many different species (Stiles 2000). It certainly seems to be present in most mammals, and its level of sophistication in any species seems to be related to the level of socialisation in that species, not just cognitive capacity (Stevens and King 2012; Borrego and Gaines 2016). Knowing how others in your social group are likely to react to you allows you to plan your day, avoiding enemies and staying close to friends. It also allows you to map your social environment, to understand who is on your side and who is against you. However, the modeller themself does not need to be attentionally present as an entity in their own modelling – they are the unacknowledged constant with which all relationships are formed, the invisible hub around which their social model revolves.

How to make models of relationships between others

Modelling others in relation to an unmodelled self requires a simple social arithmetic, in which friends count as positives and foes as negatives. In this first-level social modelling, the individual maintains a cognitive register of their relationships with other members of their group. It is a simple and direct mechanism whereby the emotion the individual feels toward another group member is their relationship with that group member; so this type of social modelling does not need to be attentional. We can describe this as one-argument Relationship-A modelling, because the modeller needs to identify another individual (the ‘argument’, A) and remember their emotional relationship to that individual.

Relationship-A modelling is not uncommon in nature, but it is not the only type of social modelling we see. A second, more complex, level is also present in primates and some other animals. This second-level social modelling works because it enhances the knowledge that an individual has about their group, allowing them to more effectively
navigate and manipulate the relationships in that group. It is sometimes equated with Machiavellian Intelligence (Whiten and Byrne 1988), although that is an extension of Relationship-A modelling in a different direction: the Machiavellian individual is able to use their relationships with others against those others. It is characterised by strategic alliances and vendettas, misdirection and misinformation, and a lack of vigilant sharing and other social rules for alpha suppression.

Second-level social modelling is quite different from first-level: it involves the modelling of relationships between others, which we can describe as two-argument A-Relationship-B modelling. The modeller’s feelings toward each of the modelled others are not informative about the emotional relationship between those others; in fact, the modeller’s own feelings can interfere with effective modelling. So, to properly understand an A-Relationship-B model, the modeller has to ignore their own emotional relationships with A and B. As well as a more complex social grammar, the modeller needs a more complex emotional grammar – they need to understand their emotions as meta-references, where understanding is not the same as currently experiencing the emotions (Edwardes 2010, 98–9).

A-Relationship-B modelling requires a considerably more sophisticated cognitive capacity than Relationship-A modelling: the modeller has to model others, as before, but they also have to accurately model the relationships between those others. I can like, or dislike, or fear Alf, and my emotion represents and dictates my relationship with him: what I feel directly affects my behaviour. However, the relationship between Alf and Beth is not the same as my relationship with either of them, so I cannot let my own feelings toward either of them influence my understanding of their relationship with each other. For instance, my relationship with Alf is bad, while my relationship with Beth is good; but I also know that the relationship between Alf and Beth is also good; so, if I wish to ally with Beth, I need to accommodate her relationship with Alf, which is contrary to my own relationship with Alf. This social calculus may be employed by chimpanzees and bonobos as well as humans, and something like it has been observed in other primates (Whiten and van Schaik 2007), which makes the characterisation of non-human primates as unrepentant Machiavellians seem somewhat unfair. However, A-Relationship-B modelling is considerably less common in nature than the Relationship-A modelling of social arithmetic.

Does A-Relationship-B modelling also require a new type of selfhood? Probably not. This would only be needed if the two-argument form of social calculus replaced the one-argument form of social arithmetic;
but why would it do so? The new social calculus just needs to interface with the already-effective old system, it does not need to replace it. This incremental approach corresponds with what we know about evolution (Futuyma 2015) and also with what we know about social calculus in the human brain: there seem to be at least two systems of social modelling at work, one to experience affective reactions, and one to model them (Lucas et al. 2015).

A-Relationship-B modelling has implications for ToM (Baron-Cohen 1995). This is the knowledge that others have their own minds with their own agendas, and, by adapting my actions to accommodate the agendas of others, I am able to achieve my aims with less conflict than if I ignore them. ToM also seems to allow the individual to plan their own relationships with others, based on the relationships between those others and the modeller’s own relationships with them; it therefore implies a level of conscious control over the modelling process itself. In contrast, the Machiavellian intelligence of Relationship-A modelling does not need the level of awareness that we usually associate with ToM – it is all about automatic modelling of what is happening with objects ‘out there’, beyond the edge of the self, in a selfhood-free zone. Machiavellian intelligence is not about accommodating the agendas of others, it is about using the relationships of others to advantage the unmodelled self.

The existence of ToM in human social modelling implies that there must be a merging of the individual’s Relationship-A modelling with their A-Relationship-B modelling. This allows the modelling individual to begin to treat the modelled others as subjects with their own agendas, and opens the way for enhanced cooperation and joint enterprise. These, in turn, create an environment where communicative cooperation becomes useful; and this pushes human social modelling to a new, third level, where we are sharing our models of others with those others.

**Sharing models of others**

To progress from a system of internal social modelling to externally communicated modelling, ToM has to become a conscious activity. We cannot share our internal models until we are aware that we are making them; and we cannot become aware that we are modelling other individuals until we can see them fully as individuals. The automatic modelling of Machiavellian intelligence is insufficient to provide us with this awareness. However, we have only primitive methods, at present, to interrogate extinct species for signs of conscious social modelling (and
methods for detecting it in living species are not much better); so we cannot know with certainty when human ToM began. What we can do, though, is try to identify the mechanism that made social modelling a conscious activity, such that it could be communicated to others.

In fact, the processes of communication and awareness of social modelling are probably two halves of a single, ratcheting evolutionary effect. Early species of *Homo* are likely to have had primitive communication mechanisms based around vigilant sharing – see David Erdal and Andrew Whiten’s (1994) idea that everyone in a group makes sure that everyone else is not taking more than their fair share. This process is both an external expression of internal modelling and evidence to others that the vigilant individuals are using social modelling. It is not a direct, referential signal (although the sanctioning of the greedy would be), but it is inferentially communicative: other individuals become aware that social modelling is not a capacity available just to them, it is being used by those around them. Or, to put it another way, it is evidence that other individuals in the group have intentional awareness of others. If the gains that this awareness of other-awareness gives to cooperative individuals are smaller than those it gives to Machiavellian individuals, then the species will tend to become more Machiavellian; and this seems to be the path that the patriarchal chimpanzees have taken. If, however, the net gain is the other way, the species will tend to develop a more informed social cooperation; and this seems to be the path taken by the matriarchal bonobos and early humans. Vigilant sharing does not automatically lead to sharing of social models, but it can be an important impetus in that direction.

Sharing models of others using A-Relationship-B constructs requires a complex interface between social cognition and signalling cognition. Both signaller and receiver must maintain a modelled set of other individuals in their group, and a modelled set of their own relationships with those others; and they also have to keep an attentional tally of all the A-Relationship-B constructs in their group that are relevant to them. This is a very language-like system: objects (the individuals) are associated through relationships, producing propositional, or sentence-like, meta-knowledge of the social structure of the group. To exchange these models, individuals must negotiate to sound–meaning pairings, both as name-signs to represent members of the group and as representations of the relationships between the name-signs. It seems like a problem with many levels of complexity. If neither individual has sound–meaning pairings in their own internal social modelling, how are they both able to generate them? And, even if both individuals have already created their own
sound–meaning pairings, how do they merge them into a single system? Even if these two problems are solved, there remains a third: how does the single system adapt for new social group members, new relationship types, new communicators and new conventions? The process of bringing social models into communication is not a natural given, and will be discussed in more detail in Chapter 5.

Fortunately, however we got here, there is evidence that modern humans do use language to exchange social models: we, unlike all other living primates, are natural social communicators – or gossips, as Robin Dunbar (1996) expresses it. A large part of language involves exchanging social information about shared acquaintances, and sometimes about individuals known to only one of the correspondents in a dialogue. At first glance, this social information exchange appears to be highly advantageous to all parties: it creates an environment in which cheats find it difficult to prosper because their cheating can be shared by the cheated, and so become universally known by all the other group members; it allows the non-cheating group members to unite against the cheater and cooperate in sanctioning and excluding them; and it therefore rewards the virtuous and punishes transgressors. However, while this is a good description of how successful human groups organise themselves, it is also an insufficient evolutionary explanation of how we reached this position.

The problem with social communication, as Camilla Power (1998) shows, is that it is, itself, wide open to cheating. As we saw in Chapter 2, first, there is the sender’s dilemma: I know something you don’t know, which gives me an advantage over you; why should I give away that information, and thus my advantage, to you? Second is the receiver’s dilemma: if the sender has control over the information they give me, why should they give me true information? True information costs them and advantages me, while false information advantages them and costs me; so, if it is to their advantage to lie, why should I believe any volitional information they offer me?

The sender’s and receiver’s dilemmas are common to all volitional signalling; but there is one further dilemma for the receiver that is a product of the particular nature of social communication: the confirmation dilemma. When a vervet monkey gives the leopard call, it is best for other vervets to act first and ask questions later; but what happens if the call is false? Cheney and Seyfarth (1990, 213–15) provide a case in evidence involving Kitui, a low-ranking older male vervet. His age meant that he was heading down the ranks, and any new male joining the group was likely to push him down even lower. This stressful existence
may have made his signalling less reliable, as it did for Melton. Kitui was recorded giving a false leopard alarm call on three separate occasions; in each case there was no leopard, but there was an unaffiliated male vervet trying to join the group. Kitui’s call kept the new male in one tree, away from the main group in another tree. Unfortunately, Kitui did not understand that, for the deception to be fully effective, his own behaviour should correspond with his call; his failure to climb a tree himself while calling gave the game away, and the other vervets began ignoring his deceitful calls.

From this we can see that vervets’ responses to leopard calls are subject to confirmation after reaction: if they are not confirmed, then the call – or, at least, the call–caller pairing – loses its value and becomes meaningless. However, shared models of the relationships between others, unlike warning calls, are not immediately verifiable: they are mostly internalised interpretations of observed events, and rely on evidence that is usually in the past. They are also opinions, not facts, and they are heavily biased by the sender’s point of view. It is virtually impossible to confirm the truth of shared social models; so why should the receiver pay any attention to them?

There are two possible solutions to these trustworthiness dilemmas. The first is that social communication is not about explicit meanings, it is about implicit meanings: it is an attempt to engage the receiver in a social activity, in the same way that grooming does. The primary purpose of sharing a social model is to build alliances: to identify whether the receiver has the same views, and therefore the same intra-group objectives, as the sender. The sender is only superficially offering their own view of a particular social relationship; more importantly, they are actually attempting to negotiate to an agreed social calculus surrounding that relationship. When a sender offers the model ‘Alf likes Beth’, they are inviting the receiver to participate in a negotiation toward meaning in which the sender, as well as the receiver, is open to revising their social calculus.

The second solution is that social communication does not need to be truthful to be valuable. When I offer my interpretation of a social relationship, I have to make it consistent with the social calculus you already have, if it is to be believable; and the only model I have for your social calculus is my own social calculus. So when a sender offers the model ‘Alf likes Beth’, they are telling the receiver more about their own relationships with Alf and with Beth than about the actual relationship between Alf and Beth. Each individual message is like a piece of jigsaw; and the more pieces the receiver has, the better their understanding of...
the sender’s worldview (or group-view, at least) and of the network of relationships surrounding the sender. Even if a sender deliberately tries to give false information, pre-acquired social communication means that almost everything can be cross-checked; it cannot be confirmed as true or false, but a message that does not fit properly into the jigsaw is probably false.

Both of these solutions are dialogic: they work because individual signals only have value as part of a continuing social exchange. The dialogue between individuals is not comprehensible in terms of individual messages, only in terms of the whole communicative experience. The message in an utterance no longer needs to stand or fall by its correspondence with the real world. It represents only a single data point in a much larger dataset; and that dataset, in turn, relies on opinion and viewpoint, not existential truth.

Both solutions also benefit from a more complex form of social modelling, in which the receiver is able to tag each received A-Relationship-B model with the identity of its sender. This gives an A-Relationship-B-by-C form, in which the truth in a shared social model is not absolute but nuanced by what the receiver knows about the sender. This modelling is therefore hierarchical: the identity C ‘governs’ the natures of A, B and the relationship between them – their natures are real only in terms of C’s utterance. When this new form of social modelling is shared with others as a ‘C-said-A-Relationship-B’ utterance, two things happen. First, the sender gains deniability over the truth of the utterance: the sender is not saying A-Relationship-B, they are only reporting C’s utterance. Second, the receiver can tag this new message in their social calculus with the identity of the new sender, creating an A-Relationship-B-by-C-by-D form (Edwardes 2014). Sharing social models does not just enrich the receiver’s social calculus, it also introduces a new level of sophistication to modelling, one that has the potential for recursive cognition – and recursive communication.

Shakespeare uses this sharing of social calculus to great effect in *Much Ado about Nothing*, in which Benedick and Beatrice change their views of each other because of two conversations that are staged by their friends so that they overhear them: Benedick overhears that Beatrice ‘loves him with an enraged affection’, but will not admit it (Act 2, Scene 3); and Beatrice overhears ‘that Benedick loves Beatrice so entirely’, but Beatrice is ‘self-endeared’ (Act 3, Scene 1). Both decide to act upon these staged social calculus models, believing them to be true because they were overheard; and it is their resulting alliance that brings the play to a happy conclusion.
Making models of my self

Once the gossip machine is up and running, communication itself begins to introduce new features to social modelling. Social communication relies on individuals having an awareness of the intentionality of others: communicable social modelling requires an understanding that others have minds, which means that minds are possessed by both the individuals being gossiped about and the individuals being gossiped to. The process of social modelling is, by definition, the capacity to model others, which can be managed without conscious attention; but the extra complexity involved in the communication of those models means that they have to become consciously managed. Additionally, while cognitive social modelling is completely about third-person models of other people, communication introduces an extra complication: it is possible for the sender to offer their model of the receiver to the receiver. Talking to the receiver about the speaker’s model of the receiver brings to the attention of the speaker the receiver’s role as listener, and the particular interest of the listener in the model being shared. Thus, the model takes on a second-person significance for the sender. This new significance does not make the model different from that of any other third-person model, but it changes the pragmatics of the communication act. Informationally, you need to tell the listener honestly what they need to hear; but pragmatically, you also need to present the information in a way that makes the listener want to hear it. Sharing models of an individual with that individual means that the model has to be subjectively comfortable as well as objectively accurate to the listener.

If this seems like a big task, consider what happens when the listener receives a model of themself. First, they have to be able to incorporate that model into their social calculus; and the only way they can do that is to treat the model like any other received third-person model: they have to model themself as a third-person entity in their own social modelling. Second, and more significantly, what the listener has received is somebody else’s third-person model, which happens to co-identify with their own model of themself; except that, in the pre-communicative unshared state of social modelling, there was no need for a consciously available model of themself. So the received model is simultaneously the only conscious evidence that the individual has about themself, and a second-hand opinion. It is likely that, as the individual adds more third-person models to their self-model, they will begin to develop an awareness of themself as a first-person entity. However, this is an awareness of an amalgam of other people’s third-person models. We each end up with a
certainty that we have a self, but the self that we have is actually a model, and not a true awareness of a real self. We are back to the position of Metzinger, Wegner, Nørretranders and Hood (see Chapter 1), that there is no actual self behind the modelled self that we insist is us. So when we make a model of ourself, do we not see anybody, as Alice believes? Or do we see nobody, as Alice says?

This awareness of self, regardless of how inaccurate or irreal it may be, does generate an important new cognitive awareness: that there is a personal self to be had. Awareness of the modelled self, and awareness that it is a model, generate the awareness that I have selfness—a me-ness, a myself-ness and an I-ness.

Me, myself and I

As we saw in Chapter 1, Freud asserted that we all have three selves: the id, a largely subliminal self that enacts all the basic instincts of being human; the super-ego, the conscious self which enacts all the social and cultural activity of trying to be the best human we can be; and the ego, which enacts all the aware cognitive activity of being human, and which tries to choose the best path between the demands of the id and the aspirations of the super-ego. For Freud, none of these were modelled selves, they were all real; but, while the id clearly has no awareness of itself, the super-ego seems to be aware of itself as a social entity, while the ego is aware of all three selves. We now recognise that this image of the individual at war with itself is a useful metaphor for many psychological disorders, but it does not reflect the reality of the psychologically well-ordered person. It is little wonder that this definition of selfhood led Freudian psychologists to see psychological disorder everywhere.

Language, like Freud, gives us three views of our selfhood: the subjective I, which instigates activity; the objective me, the recipient of activity; and the reflexive myself, the me as visualised by the I. This linguistic representation also reflects how our self-modelling works: the subjective I makes a model of the objective me to generate the reflexive myself. We may feel that, like Freud’s id, the subjective I provides a subliminal basis for selfhood; but the linguistic view means we can also step outside the model to view all three as irreal, consciously represented models, with the objective me being a model made by a model, and the reflexive myself being a model of a model made by a model. As with all words in our language, the first-person pronouns me, myself and I are
metaphors for realities. Usually, the realities represented by words are external to the person generating the words; but in the case of first-person pronouns they appear to be internal realities. This, however, is just another feature of the selfhood illusion: first-person pronouns are functionally external in that they are models of internal reality projected onto the real world. Modelling is always a cognitively internal activity, but its products are externalised; which is what allows us to see our models as models of something.

Our relationship with our pronouns is complex. For Émile Benveniste (1970 [1996]), there is an important attentional difference between the different persons: the third person represents an object outside the communicative act, so is fundamentally different from the first and second persons, which are inside the communicative act; and the first person itself creates a subjectivity in language, allowing the speaker to simultaneously represent themself as the subject, or ego, of an utterance and the producer of it (Benveniste 1958 [1971]). It also seems that there is a significant difference in usage between me and myself, at least in terms of what is being modelled. Usages of me and myself in the constructs I love/like/dislike/hate me/myself seem to reflect the distance of the model from the self: me is, relative to myself, more loved and liked than disliked and hated by I, and more loved and liked in the present tense than in the past (Edwardes 2003). James Pennebaker (2011) shows that the way we use our pronouns (and other small function words) can also be a window onto our personality and intentions. For instance, how frequently we use first-person pronouns is an indicator of how formal we are being – in the terminology used here, how formal we want our relationship with our conversation-partner to be. In formal discourse, we model our own self more as a remote myself and less as a familiar I, and we model our conversation-partner as they rather than you. The whole conversational model has been moved a step back from intimacy by depersonalising the models of the speaker and listener. The role of pronouns in selfhood will be further explored in Chapter 6.

These are just some of the ways in which language is both the progenitor of conscious selfness, and acts to define how we model our own selfness and the selfhood of others. Selfness exists because we began to share our cognitive models of others, and then had to deal with the cognitive complexity that sharing produced. Me, myself and I are the products of a highly cooperative communication strategy overlaid on a sophisticated pre-existing other-modelling capacity.
Awareness of selfness: for humans only?

The fortuitous confluence of events that led to awareness of selfness in our species raises an important but often glossed-over question: why has it only happened to humans? The answers we can give to this question are helpful in defining us as a species, and in understanding the nature of our differences from other species. So what answers are we able to give?

It turns out that the view of language as a Great Leap Forward in evolutionary terms (Smith and Szathmáry 1999) may not be as justifiable as we once believed. Language is not the pot of gold at the end of the cognitive developmental rainbow; instead, like all species-specific traits, it opens up new ways of being by closing down other ways that still work effectively for other species. The cognitive developments that led to language and awareness of selfness bring problems as well as solving them.

The first big disadvantage lies in cognitive development itself: as discussed in Chapter 2, having a big brain is incredibly costly, consuming about 20% of the human energy budget (which means that every fifth doughnut goes straight to the brain and not to the hips, as some dietary pundits would have us believe). This disproportionate energy consumption is not caused by complex or conscious thinking, however; our brain's energy usage seems to remain the same whether we are awake or asleep, solving quadratic equations or daydreaming, modelling complex interpersonal relationships or gazing at clouds. The energy consumption is not caused by how we think or what we are thinking about, it is simply a product of maintaining the big brain itself; it is the Red Queen, running as fast as she can just to stay still.

Yet the role that this big, expensive brain plays in our species remains opaque. Clearly there are some types of complex thinking that our species needs to do; but what are they? If we look at many other large-brained species (such as cetaceans, other apes or elephants) we find that, like us, they are highly social; so is the evolution of big brains associated with social living in large groups? Robin Dunbar (2010) thinks this is the case, and points to a correlation between brain size and group size, a correlation that seems to approximately work for most primates. However, the correlation breaks down when applied as a general evolutionary rule: the most social animals on this planet are eusocial insects (ants, bees, wasps and termites), which can live in communities numbering tens of thousands; yet their individual brains are tiny. At the other extreme there are many species of octopi: clever, problem-solving animals with relatively large brains but no social life to speak of. Other than territorial aggression, the only same-species communicative act
in the life of an octopus is an act of coitus, which is followed by death for both sexes – within weeks for males and within months for females (Godfrey-Smith 2016, Chapter 8). So whether group size is proposed as the driver for brain size or brain size as the driver for group size, there seem to be significant examples in nature arguing against both positions. Additionally, while Dunbar predicts a human group size of 150 based on brain size, human joint enterprises can involve much larger groups without clearly demonstrating the group fission that Dunbar predicts. It seems that, somehow, humans have sidestepped the group-size limits, just as they seem to have circumvented the natural limits on brain size and communication complexity.

The products of the large brain are even harder to justify. We have already seen the communicative problems involved in a volitional communication system like language: the sender’s problem of sharing information that is actually most valuable to them when only they know it; and the receiver’s problem of trusting information that is most valuable to the sender when it is a lie. When a communication system becomes volitional, it should collapse into meaninglessness, not transform the species’ socio-cultural organisation.

These problems are insignificant, though, compared to the socialisation problem: if bigger brains make for bigger groups, or vice versa, what is the fitness mechanism that favours bigger groups? Large groups face more challenges than just the socio-cognitive need to keep track of more individuals; there are the problems of environmental load-bearing, fission–fusion and shared enterprise, all of which also have to be overcome. The environmental load-bearing problems are: first, that larger groups will exhaust local resources much quicker than smaller groups, so they need to be more mobile than smaller groups; second, that larger groups will exhaust larger territories, so will have to move further to escape their own depredations; and third, that a species with a small number of large groups is more vulnerable than a species with a large number of small groups to epidemics that wipe out whole groups. The fission–fusion problems are: when a group exceeds its maximum size and has to split, there have to be enough individuals in both new groups to ensure their survival; and when two small sub-optimal groups meet, they need to have strategies to help them join together to make a more optimal group. The shared enterprise problem is: the larger the group, the more enterprises there will be to be shared; but how are the individual needs and skills matched efficiently? All of these aspects of the socialisation problem can be mitigated by effective communication; but that only weaves the communication problems into the socialisation
problem, making the co-evolution of large brains and large groups even more perplexing.

So if awareness of selfness relies on language, and language is a product of brain size and group size dynamics, and human brain size and group size are both difficult to explain using evolutionary calculus, then we should not be surprised if only humans have an awareness of selfness. If, in addition, we can show that an awareness of selfness is, itself, problematic in evolutionary terms, then the question changes from ‘why do only humans have awareness of selfness?’ to ‘why does any species have awareness of selfness?’ The answer to this may lie in the development of another product of language, this time from the non-personal meanings generated by negotiation: human culture.

**Language, culture and the self**

Human culture, like human communication, is simultaneously continuous with the rest of nature and quite distinct from the cultures of other species: it has a different approach to what counts as information – and, therefore, what counts as knowledge. Donald Brown (2004) describes human knowledge as consisting of two types of information: *etic facts*, which are true\(^1\) by their nature but not necessarily consciously appreciated (such as ‘a tapir usually has four legs; roasted tapir is good food’); and *emic facts*, which are true because we agree they are true (such as ‘the tapir is the national animal of Belize; it is wrong to kill and eat tapir because they are endangered’). Etic facts are common throughout nature and, if they enhance individual fitness, can become genetic facts (for example, the fact that snakes bite and can kill has been encoded into our genome, justifying a healthy anxiousness around snakelike objects). They should not be confused with mind-independent facts: etic facts are etic because they reflect the external world, but they are facts because they are ideas commonly shared among humans.

As well as becoming genetic facts, however, etic facts can also be conventionalised as cultural facts. For instance, a link has been established in chimpanzees between self-medication with rough leaves swallowed whole and the increased expulsion of gut worms. Desmodium leaves are used by the chimpanzee group at Gashaka, Nigeria, but other leaves are used by different groups elsewhere (Fowler et al. 2007). This behaviour is cultural: different chimpanzee groups have found similar, but not identical, solutions to the same problem. However, it is also a cultural solution based on etic facts; and, because it is based on such
facts, it actually works. Compare this to facts that have informed recent human medical culture: in the late 1800s, cigarette smoking was touted in Europe as a treatment for asthma, on the basis that inhaling smoke dried up the excessive mucus thought to cause asthma (Jackson 2010); mercury remained the main treatment for syphilis throughout the 1800s, until it was replaced by arsenic in 1910, and, finally, penicillin in 1943 (Frith 2012); and, even today, the biggest threat to tigers across the world is the trade in animal parts for ‘medicine’ (Byard 2016). All of these are emic medical facts, based on local cultural beliefs that have no basis in etic facts, or which rely on coincidences of success as evidence.

Etic facts, therefore, are not the same as natural facts, and they can include cultural facts; what they all share in common is that they do not rely on a belief system to enforce them, which means they have to be intrinsically valuable to the knower. This is in contrast to emic facts, whose worth lies in the solidarity they create in the group: they are extrinsically valuable to the knower because they help the knower navigate their society by conforming to its culture. A culture based on etic facts is qualitatively different from an emic culture; but, because etic facts do not need to be genetically encoded, non-human cultures based on etic facts can, and do, exist. However, emic facts form the basis of most human cultures. Humans are outstandingly good at creating and enforcing emic facts, basing them on agreement rather than evidence. This becomes unsurprising if we accept the proposition that human communication is itself based on emic facts: our languages work not because there is a special ‘language mechanism’ inside each of us, but because we are able to negotiate toward meaning. This willingness may well be genetic, but the negotiation itself requires the ability to consciously accept another person’s imaginings as valid on some level. As Lewis Carroll puts it:

‘But she must have a prize herself, you know,’ said the Mouse.
‘Of course,’ the Dodo replied very gravely. ‘What else have you got in your pocket?’ he went on, turning to Alice.
‘Only a thimble,’ said Alice sadly.
‘Hand it over here,’ said the Dodo.
Then they all crowded round her once more, while the Dodo solemnly presented the thimble, saying ‘We beg your acceptance of this elegant thimble’; and, when it had finished this short speech, they all cheered.
Alice thought the whole thing very absurd, but they all looked so grave that she did not dare to laugh; and, as she could not think
of anything to say, she simply bowed, and took the thimble, looking as solemn as she could.

(Lewis Carroll 1865, Chapter 3: ‘A Caucus-Race and a Long Tale’)

Like language, exchanging social models also requires negotiation toward meaning: we each accept and use the unverified models of the social relationships of others when they are offered to us, even though we know them to be emic opinions and not etic facts. This acceptance of the opinions and beliefs of others about others unlocks all kinds of useful linguistic and modelling tricks and devices, such as referencing non-current events (temporality), referencing possible but not yet actual events (modality), and using shared imagination. It also has an effect on how we model ourselves. Social modelling gives me access to what other people think (or say they think) about me, allowing me to build a model of myself as a social being. This social self-model is an emic fact, a third-person representation of my self as an entity in my social calculus; but I can treat it as an etic model inasmuch as it represents an objective view of me as an other; it’s the best understanding of my self available to me.

While human culture is an outcome of human socialisation and language, it nonetheless generates yet another, and very different, self-model. With its emphasis on emic facts, human culture presents me with an ideal model of what an individual should be in the particular culture in which I find myself. It is an aspirational self-model – still external to (and different from) the Actual self (which remains unknowable by direct introspective methods), but also different from the third-person self-model provided by the sharing of social calculus (or social communication). The Cultural self is based on the emic social expectations of others rather than their mostly emic social knowledge – and, as I am a member of the same culture, they are probably expectations that I (my social self-model) have about myself.

This emphasis on emic facts in human culture creates a very odd inversion in the social strategies of our species. Like eusocial animals, we have societies with high levels of organisation, complexity, cooperation, individual specialisation, task-sharing and self-sacrifice; but where the eusocial lifestyle involves a physical culture bound by genetic imperatives, human society is governed by symbolic culture. Eusocial societies work because of the high level of relatedness in a nest and the fact that there are few fertile females – usually only one per nest. The only way for the sterile nest members to get their genes into the future is to protect the queen, their mother, and her fertile offspring, their sisters
and brothers. This means that the range of cultures possible is severely limited, because they have to be based on etic realities, not emic beliefs. In contrast, the high levels of organisation, complexity, cooperation, individual specialisation, task-sharing and self-sacrifice in human cultures are all generated emically, through group expectations and shared beliefs. The correspondences between the needs of eusocial and human societies help to explain why humans seem to have adopted a pseudo-eusocial social system; but, where reliance on etic facts makes the cultural range available to eusocial animals extremely small, the human range of cultures, based on emic facts, is bewilderingly large: any set of shared beliefs can become the basis for a culture.

An outcome of relying on emic facts is that, whereas eusocial cultural systems are stable and durable, human cultures are vulnerable to collapse and elimination when key beliefs are challenged. There is little durability in human cultures, which tend to last only hundreds of years rather than the millions of years for eusocial animals; but this lack of durability does have a surprising genetic effect. The constant turnover of cultures means that there is a process of succession: cultures that place a greater reliance on etic facts have a survival advantage over more emic cultures. History has shown us that cultures with greater reliance on etic facts tend to win in any competition against cultures relying on emic facts (Diamond 2005). So any genes that favour reliance on etic over emic facts become more common in etic human cultures, and the greater survival rate of etic cultures spreads those genes across the species.

One example of this process is the spread of lactose tolerance. Ingram et al. (2009) show that 13,000 years ago there was no significant ability to digest cow's milk in any adult human population. Children produce an enzyme, lactase, which lets them drink milk (from a range of mammals, not just other humans), but production of lactase stopped at puberty. However, the domestication of cattle about 12,500 years ago led to a rise in digestive tolerance of lactose, the indigestible factor in cow milk: genetic changes extended the production of lactase through puberty and beyond. The domestication of cattle happened piecemeal, as a series of local events; and the abandonment of hunting and gathering in favour of pastoralism occurred over a very few thousand years. We can therefore say that cattle domestication was a cultural event; but it enhanced group survival by providing a reliable and regular food source, so it was also an etic event. Lactose tolerance spread as an etic fact that favoured domestication, but it remains a cultural event: adult intolerance remains high in some Asian populations.
where milk did not become a food staple; and, across the species, two thirds of adult humans are still lactose intolerant.

Unlike lactose tolerance, however, the ideal Cultural self is a completely emic fact: it has no existence outside the minds of the members of the culture. Yet it does have one important effect on the individual: being aspirational, the cultural self-model seems to have worth as a shortcut to social self-modelling. Instead of trying to cobble together a composite self-model from the opinions about me provided by others, I can simply try to conform to the self-model that others around me treat as a cultural template for being an ideal human. A series of children’s books, *The Best Me I Can Be* (Parker 2007), indicates that, far from being a poor alternative to social self-modelling, cultural self-modelling is a common view of how effective models of self are produced, and is commonly used around children. As we will find out in Chapter 4, cultural self-modelling may be an effective way to activate self-modelling in children before they develop the capacity for effective social self-modelling. The existence of fairy stories, moral tales and mythic systems in almost all human cultures indicates that sharing exemplars of cultural morality is common, while the recorded age of some of these stories indicates that they represent an ancient human tradition.

However, the emic nature of the cultural facts in these stories means that their worth is negotiable, unlike the etic social facts that social self-modelling provides. The cultural self-model is not a product of how I am seen or how I see myself; it reflects how I believe I should be. It does not provide accuracy, but it does provide acceptability.

One final feature of the cultural self-model should be highlighted: the reasons why self-modelling is a fit strategy today are different from the reasons why it was a fit strategy when it first appeared. Ernst Haeckel (1866) is famous for the statement that ‘ontogeny recapitulates phylogeny’, which encapsulates his belief that every human embryo goes through the same developmental forms that produced our species, from single-celled creatures through intermediate species-like forms up to modern humans. This idea, known as Recapitulation Theory, has been rejected since the middle of the twentieth century as a plausible genetic explanation (Blechschmidt 1977, 32). It also seems to be unworkable as an explanation for self-modelling. At the species level, cultural self-modelling only became possible because sharing social models led to social self-modelling; and this, in turn, established a generalised cognitive potential to self-model. Human culture, with its emic belief in ideals, then created the shared ideal self, which the individual could
then own as a cultural self-model. In contrast, at the individual level, the human child initially seems to adopt the cultural self-model to define themselves, only later replacing it with their own personal self-model, which they build from their social calculus exchanges. Ontogenically, at the level of the individual, we seem to start with the generic ideal cultural self-model because it is cognitively simpler – even though, phylogenically, at the level of the species, it is an emergent feature of social calculus and language.

The disadvantages of a modelled self: deficient self and self-deception

It seems to be a common belief that the capacity to accurately model yourself is a Good Thing. The self-modelling individual is able to map themself into their social calculus accurately and mindfully, allowing them to manipulate their group in ways that less enlightened individuals (perhaps those still relying on a cultural self-model rather than a social self-model) cannot manage. Indeed, human history is full of comments promoting and praising this level of self-awareness: the Oracle at Delphi had the maxim ‘Know thyself’ over its entrance; the fifth-century BCE philosopher Lao Tzu, the founder of Taoism, said ‘He who knows others is wise; he who knows himself is enlightened’; and Pythagoras said ‘No one is free who has not obtained the empire of himself’, to which Socrates added, ‘True wisdom comes when we know how little we know about life, ourselves, and the world around us’. In Hindu doctrine, knowing your eternal self, or atman, is the route to enlightenment; and in more modern times, Robert Burns said ‘O wad some Pow’r the giftie gie us, to see oursels as ithers see us!'; while Oscar Wilde advised against taking the cultural-self shortcut to self-knowledge, with: ‘Be yourself; everyone else is taken’.

Yet, as we have seen, the only conscious representation of our self that we have available is what we have cobbled together from the social and cultural models offered to us by others. When we model our self we are modelling ourself from the outside looking in; but our unmodelled selfhood is imposed on the world from the inside looking out. The self-knowledge that language allows us is not reflexive but reflective, creating an image of our self that is recognisable and acceptable to others, and which we can then use to define and refine our self-model. This externalised vision of our self is, therefore, not so much what we are but
what others believe and want us to be; it is a model of the socialised and enculturated self, a representation we advertise or aim for rather than actually are.

Social and cultural self-modelling means that we are constantly trying to meet expectations imposed on us by others: human socialisation requires us to see self-promotion and hubris as vices, while human culture promotes humility, self-effacement and modesty as virtues. We also cannot avoid comparing our own third-person modelled selfhood with the other third-person models in our social calculus, and with the ideal self that symbolic culture imposes on us – all of which means that we are always finding ourselves wanting. Indeed, the human cultural view, at least in the West, seems to be that everyone is incomplete and improvable: whatever we are, we could be better.

Several apparently unfit evolutionary strategies are generated by our self-modelling: conversational and social turn-taking, generosity to strangers, acquiescence to group norms and so on. These are all strategies that, in a Machiavellian society, disadvantage the individual by allowing others to exploit their naivety. However, these individually unfit strategies are precisely what make us successful as a pseudo-eusocial species, as they encourage high levels of organisation, complexity, cooperation, individual specialisation, task-sharing and self-sacrifice.

Our self-models even allow us to deceive ourselves about the personal value of self-sacrifice – which remains, in all cases except very occasional kin-saving actions (Hamilton 1964), inexplicable in selfish gene terms (Dawkins 1989). Honoré de Balzac (1835 [1991], 125) identified one reason why this may be so, when his character Père Goriot said, ‘Some day you will find out that there is far more happiness in another’s happiness than in your own’: it seems that self-sacrifice has become so common that we may even get a genetically directed emotional reward from it. It is certainly encoded in our culture, where emic ‘evidence’ offers us rewards beyond the physical. At the end of Charles Dickens’ *A Tale of Two Cities* (1859 [2000]), Sidney Carton saves Charles Darnay from the guillotine by taking his place. His final utterance, ‘It is a far, far better thing that I do than I have ever done; it is a far, far better rest that I go to than I have ever known’, summarises the two great emic rewards for self-sacrifice: reputation and survival after death.

It seems, therefore, that awareness of selfness – like human-sized brains or human-sized groups or human language – poses yet another evolutionary conundrum. How has a species become so unusual and apparently evolutionarily unfit without going extinct? And why, despite all these unfit aspects of our species, have we become so dominant on
this planet? These are questions that will be tackled in Chapter 5; but we will first take a step back from the evolutionary improbability of our species and look at how we turn small humans into adult humans. We will review the innate mechanisms that can be activated to make us conventionally human, and the external mechanisms that we impose on trainee humans to ensure their assimilation into human societies. It is time for the nature–nurture debate.

Notes
1. The word ‘fact’ is used for emic and etic facts because they are not true or false in a conventionally logical way. For instance, ‘elm trees are taller than oaks’ is a logical proposition, capable of being proved or disproved from evidence; ‘this bonsai tree is tall’ is an etic fact, if particular definitions of the words ‘tree’ and ‘tall’ are accepted; ‘tall trees are better’ is a value judgement, which is true if people agree it is true.
2. Eusocial in its cooperation, but not eusocial in its reproduction.