‘Come, there’s no use in crying like that!’ said Alice to herself, rather sharply; ‘I advise you to leave off this minute!’ She generally gave herself very good advice (though she very seldom followed it), and sometimes she scolded herself so severely as to bring tears into her eyes; and once she remembered trying to box her own ears for having cheated herself in a game of croquet she was playing against herself, for this curious child was very fond of pretending to be two people. ‘But it’s no use now,’ thought poor Alice, ‘to pretend to be two people! Why, there’s hardly enough of me left to make one respectable person!’

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

In 1871, Charles Darwin published *The Descent of Man, and Selection in Relation to Sex*, the second book of his evolutionary duology. In this book, he set out to show that humans, with all their cultural accoutrements, descended from ape ancestors. He took the view that we should expect no great novelty in human cultural tools – including language, despite its apparent difference from other communication systems in nature. Darwin proposed a series of routes from natural sound-making to meaningful language-making: imitation, both of the sounds of other animals and of things, and by replication of physical actions (such as waving), but using the tongue and mouth; emotional sounds; and the sounds of work and play (Darwin 1874). Jespersen (1922, 114–17) later dismissed these as insufficient explanations, and they have become known as the bow-wow, ding-dong, ta-ta, pooh-pooh, yo-he-ho and la-la theories of language origins. They all remain possible routes to language, but none has yet been evidenced beyond the level of hypothesis.

Charles Hockett (1960) made the first modern attempt to identify what made human language different from other communication
systems, and he devised a set of 13 (later 16) design features of communication. The first 13 features were:

- A vocal-auditory channel, or speaking and listening (1); broadcast transmission and directional reception (2); and transitoriness or rapid fading (3): we now recognise that these are useful but not necessary for language, or for any communication system.
- Interchangeability (4): we can receive as well as transmit.
- Total feedback (5); and specialisation (6): speakers can control their speech because language is communicative and intentional.
- Semanticity (7); and arbitrariness (8): sounds have meanings, but the sound–meaning correspondence is arbitrary.
- Discreteness (9): small differences in sound can represent big differences in meaning, and vice versa.
- Displacement (10): speech can refer to non-present events, treating the irreal as real.
- Productivity (11): language allows the creation of novel utterances to represent new ideas.
- Traditional transmission (12): language is a negotiated convention; language and culture are intertwined.
- Duality of patterning (13): meaningful messages are made up of several distinct meaningful units (words and morphemes), which themselves are made up of distinct but meaningless units (phonemes).

The last three design features were added by Hockett in 1963:

- Prevarication (14): language can be used to deceive.
- Reflexiveness (15): language can be used to talk about language.
- Learnability (16): language is teachable and learnable.

Of these, Hockett initially believed that displacement, productivity, traditional transmission and duality of pattern were exclusive to language. He later added prevarication, reflexiveness and learnability to this exclusive list, but then removed displacement, traditional transmission and prevarication. Unfortunately, this attempt to identify the specialness of language as a communication system relies on a belief that there is a specialness to be identified; this remains to be proved and, indeed, may not be the case. Of the four design features Hockett initially reserved to language, we now know that honeybee communication uses displacement (von Frisch 1973), although not productivity, traditional transmission or...
duality of patterning; but many primates are now known to have local cultures within their species, and they use traditional transmission to pass on their cultural memes (Horner et al. 2006). Looking at the other two original ‘exclusive to humans’ features, chimpanzees who have been taught to communicate with humans using a hybrid gesture system have shown spontaneous productivity in their signals (Savage-Rumbaugh and Lewin 1994), while prairie dogs (Slobodchikoff 2002) and marmosets (Pomberger et al. 2018) have been found to use duality of patterning. Dolphin communication may use all four – at this stage we do not have enough evidence to know, but the indicators are positive (Herman and Uyeyama 1999). Of the final three additions, prevarication (deception) is common throughout nature (for example, Kirkpatrick 2011), and osmotic learning is common in mammals (Crockford et al. 2004). This leaves reflexiveness, the status of which is unknowable because it is ill-defined – we cannot even say whether our own capacity to use language to talk about language is truly reflexive or just a subset of modelling: when we appear to be using language to talk about language, are we really referring to language, or are we just making a model of language that we can talk about? The belief in grammar as a fixed rule system behind our utterances – among some linguists, at least – indicates this may well be so.

Wacewicz and Żywczyński (2014) indicate that Hockett’s problems with his design features stem from his treatment of language as an identifiable ‘thing’, with the design features as components or subsystems of the ‘thing’. This misses the point that language, like any communication system, is a cognitive tool for achieving objectives, and it only remains an effective system while it continues to achieve those objectives. Just as a description of a screwdriver tells you nothing about the vital role of screws in woodworking, so a reductive description of language ignores its most important function: it is a way of doing things, and the things that get done are more important than how they are done.

Daniel Everett (2017) adds a 17th entry to Hockett’s list – or, perhaps, it would be fairer to say that Everett’s single characteristic underlies the other 16. Everett calls this characteristic:

‘underdeterminacy’ – saying less than what is intended to be communicated and leaving the unspoken assumptions to be figured out by the hearer in some way. Underdeterminacy has always been part of language.

(Everett 2017, 3)
Underdeterminacy is a term he borrowed from Pragmatics (for example, Carston 2002, Chapter 1), but he uses it to make the case that language is not about what is said, but about what is meant. Any story about the origins of language has to accommodate the fact that language communication is not a series of unconnected utterances, but a series of ongoing relationships between people; and, as the relationships develop, they become more interpersonal and contextual – and the communication in the relationship becomes more underdetermined. Memory does not just remember the past; it informs the present and directs the future.

What does this tell us about the role of self in language? So far in this book, the case has been made that awareness of self is not a necessary outcome of large or complex brains; and it is not even a necessary outcome of complex social cognition. It is, however, a necessary outcome of the communication of complex social cognition. We don’t have a need for self-awareness until we are sharing models of other people’s selves, and only then when others are sharing their models of our own self with us. If my self-awareness emerges from my capacity to make a third-person model of myself, and my only need for a capacity to self-model is to incorporate your models of me into my social calculus, and you are only able to present your models of me to me because we share social calculus, then we would expect that the mechanism through which we share social calculus – a communication system that can reliably transmit the complexity of social calculus – will be replete with indicators of that complexity.

It is reasonable to assume, in the absence of a communication system specialised for social calculus, that we share social calculus through language. This is not to say that the sole function of language is the sharing of social calculus, or even that this was the original function of the precursor to language. The sharing of interpersonal information does require both cognitive complexity and a signalling system that can be segmented, differentiated and hierarchical; so social calculus certainly appears to be a strong contender for the origin of complex human language; but that claim is not made here. All we are looking for here is evidence of a close relationship between social calculus and language: the link between social calculus and language needs to be evident in the way the two of them work together.

So is there a relationship between language and the self- and other-modelling of social calculus? Alice’s conversation with herself indicates that there does seem to be. She shows a willingness to see herself as both speaker and listener in her monologue (‘I advise you to leave off this minute!’); she models herself as both a cultural she to advise and
a social *herself* to be advised; she further distances these two selves by giving the advised Social self the self-will to ignore the advising Cultural self, and the advising self the right of punishment over the self-willed advised self. Alice’s relationship with her models of her selves is simultaneously intensely human and – in the form expressed by Carroll – deeply strange. Her final utterance (‘there’s hardly enough of me left to make one respectable person!’) reflects the dilemma of self-modelling: the modelled Cultural self is not a real self, but it is the best approximation we have to the respectable self we hope others believe us to be.

**Pronominalisation and selfhood**

One of the biggest effects of social modelling on language is likely to be the existence and nature of pronouns – the words that represent the speaker, the listener and anyone and anything else referenced. Dictionaries tell us that pronouns are substitutes for nouns and have very general reference – they are not direct references to things or people, they refer to the communicative roles undertaken by things and people. But as we saw in Chapter 3, they are not merely a simplification of how we name things: they act as ad hoc labels in a discourse, allowing the interlocutors to reduce the utterance load at the cost of increased cognitive load. For instance, when we hear ‘you shouldn’t do that’, we engage in a fast comparison of the possible members of the group *you*, their current activity, the cultural expectation about that activity, and the intention of the speaker themself. Our reaction will be different depending on whether we are likely to be in or out of the group *you*, whether we feel our current action is or is not culturally acceptable, and whether we accept or reject the approbation of the speaker. The two pronouns *you* and *that* indicate people and actions only indirectly, and their underdeterminacy means that the utterance is not just context-specific, it is also listener-specific: different listeners hearing the same utterance will have different objects in mind as *you* and *that*, and thus react to the utterance differently.

This situation is further complicated by the fact that pronouns are not a stable class across languages. If we look at the French version of the utterance, we find an immediate difference. Both ‘*tu ne devrais pas faire ça*’ and ‘*vous ne devriez pas faire ça*’ are valid translations, but they do not mean the same thing: where English ‘*you*’ can refer to a single listener or several, French divides the pronoun into singular and plural forms. In Spanish, a different problem arises: ‘*no deberías hacer eso*’ and ‘*no deberíais hacer eso*’ both retain the singular–plural distinction, but
the word for ‘you’ has disappeared. It is only indicated by the ending of the verb ‘should’. Spanish is called a pro-drop language because the pronouns are not obligatory, and are usually used only for emphasis (‘tú no deberías hacer eso’, ‘you [as a particular individual] shouldn’t do that’). However, as an example of underdeterminacy developing out of familiarity, the construct ‘shouldn’t do that’ is also a fully acceptable English form. The term ‘pro-drop’ does not so much define a language as a way of using the language.

These two examples explore only a fraction of the differences between second-person reference in English, French and Spanish; and other languages add further complications to pronominal reference. Japanese is considered by some linguists to lack full pronouns, using noun phrases instead. For instance, a man often refers to himself in the first person using the word boku, which actually means male servant. This removes reference to the speaker from the utterance, turning it into a third-person reference: ‘The servant is sorry’ rather than ‘I am sorry’. Watashi is used by both genders, and means something like ‘the private self’. This referencing of the self in the third person is not unknown in English, and it is often done using the person’s name. For instance, Donald Trump’s first-ever tweet in 2009 was ‘Be sure to tune in and watch Donald Trump on Late Night with David Letterman as he presents the Top Ten List tonight!’ We even have a word, illeism, to describe self-referencing by name; but we tend to view it as either childish (under-4s commonly refer to themselves by name) or as narcissistic and somehow dishonest. In Japanese, it is seen as polite self-effacement.

Malay is another language in which full pronouns seem to be absent, and others are referred to by their social role. A Malay-speaker has no need for I or you, because they always have a role they can use. For instance, when speaking with a grandparent, the grandparent is nenda to both speaker and listener, and the grandchild is cucunda. Pirahã, the language documented by Daniel Everett, ‘has the simplest pronoun inventory known, and evidence suggests that its entire pronominal inventory may have been borrowed’ (Everett 2005, 622). Everett describes it as having only three pronouns, for the three persons, and no differentiation between singular or plural. The pronouns act as prefixes to verbs, although they can be used as stand-alone emphatics, too.

To be more accurate when talking about pronouns, therefore, we should refer not to pronouns but to a process of pronominalisation, which can be defined as reference using communicative roles rather than names or titles. The fact that a language has a mechanism for pronominalisation is more important than how that mechanism works. Nonetheless, I will
attempt here to describe pronominalisation through the lens of English. It is not a perfect representation, but it is one that all readers of this version of the text should be able to understand.

Where names come from

Before we start on pronominal replacements for naming, we should perhaps consider how naming itself developed. It seems natural that we all treat ourselves as named individuals, and that somehow our name acts as a proxy for our self in our dealings with others; but that is all evidence after the fact of naming. The truth is that, in important ways, our name is not our own. Most importantly, it is not generated by us but given to us by others. This may not be the case with dolphins, which seem to create their own signature whistles to identify themselves to others. When a dolphin uses another’s signature whistle to attract their attention, they use a variant of the whistle—thereby identifying the call as a you reference rather than an I reference. We cannot know for certain whether this is what is really going on inside the dolphin’s head, but it does seem likely that the variations in the signature whistle are there to indicate to other dolphins whether the whistle is self-referring or other-referring (King et al. 2013).

Dolphins, unlike humans, seem to generate their own name-labels: they have the advantage of being born with functional vocal equipment, and usually they have generated their own signature whistle within two months or less of their birth (Tyack 1997). They are also precocious at visual self-recognition, passing the Gallup mirror test at age 7 months (Morrison and Reiss 2018). In contrast, human infants are born with underdeveloped vocal equipment, and the neurological systems controlling vocalisation are rudimentary. By 9 months, they recognise and respond to the name given to them by the adults around them, but it is not until around 18 months that they use it back to the adults (Holinger 2012). Human infants only become recognisably self-aware, and able to pass the self-recognition mirror test, at around 18 months.

Humans do the exact opposite of dolphins. We do not develop our own name; we expect our parents, or the people caring for us, to label us soon after we are born. Recently, the apparatus of the state has also insisted that this name be registered promptly after birth, further emphasising the role of others in naming a human person. This registration requirement remains the case, despite the fact that the state itself now relies on exclusive alpha-numeric combinations, and not our names,
to label us – and, of course, these alpha-numeric state labels are also given to us and not chosen by us.

Having been given a name by those around us, we then continue to receive different names throughout our lives. Some are role-specific, such as grandparent and grandchild; some are arbitrary labels, like prisoner KJ4609; and some are comments on our appearance or nature; but they all share the fact that they are given rather than self-selected. Our own attempts to re-label ourselves usually only work if they create a believable new persona for a new audience. For some people, this name-change is a vital part of their self-definition (and may be seen as social recognition of society’s definitional error); but in these cases, the old name is abandoned as a legitimate label for the individual. Authors, in contrast, often choose pseudonyms to distance their everyday self from their authorial self: the Reverend Charles Dodgson wrote as Lewis Carroll to preserve the gravitas required by his religious and mathematical work. Sometimes authors choose a pseudonym to differentiate their authorial selves: J.K. Rowling also writes detective books, but as Robert Galbraith. Actors may adopt new names for enduring characterisations they create, such as Barry Humphries’s alter ego Dame Edna Everage. Self-selected pseudonyms are also used extensively in selfish deception, which may be why discovered pseudonyms can generate a visceral shock reaction. However, most of the famous criminal pseudonyms are given by the media, not self-selected. Even the words ‘pseudonym’ or ‘alias’ show our approach to self-selected names: they are not real names, they are other names.

The human reliance on others to name us may be unique in nature. If selfness comes out of recognition of models of my self offered by others, then the capacity to recognise my self in their offered models is vital; and early acceptance of the social label agreed by others as referring to me – my name – becomes a vital part of self-recognition. Naming itself is not unique in nature, but nominalisation is such a fundamental part of human socialisation and communication that we need to be named at birth; we cannot wait for human individuals to offer their own names.

The internalisation of a group of sounds that others use to identify us, and the understanding that this group of sounds represents us, leaves us open to the idea that our name may be an effective proxy for us as an individual; it can be used to represent all aspects of our self. However, naming is also arbitrary: a name is not a unique way to represent an individual, as we can have many contextual names. This, in turn, allows us to accept alternative labels, both for us and for others, in terms of our roles in group activities – and these include our roles in communicative acts.
We accept general labels (you, they) as referring to us via our relationship with the speaker; and, as speaker, we offer a general label (me) to represent ourself-as-speaker.

The origin of they

Pronominalisation allows us to identify the three roles of speaker, listener and talked-about when we make an utterance; so pronouns only have a role in language when it is about sharing social models, social calculus or the relationships between people and objects. When social calculus is cognitive and uncommunicated, stable internal labels (or unshared nicknames) work perfectly well in modelling the individuals involved; and communication that is not about social relationships does not require the sender and receiver of the communication to be identified inside the communication. The only role for pronominalisation in non-social communication is to indicate that a previously referenced thing is still the object of interest; pronominalisation in this form is gestural, verbally pointing to the object of interest, and may well have come directly out of a hybrid sound-and-gesture communication system (for example, McNeill 2012). The language-out-of-gesture debate is complex and still ongoing; but it is peripheral to the discussion here, so we can safely ignore it without taking sides. We can instead appeal to other reasons why the first instance of pronominalisation is likely to have been the third person (in English, he, she, it and the singular and plural they).

The first of these reasons is that third-person pronominal reference does not rely on communication, but can be completely cognitive. The simple, linear social calculus model of A-Relationship-B is sufficient to understand the emotions involved in that single pairing; but if I am to build a cognitive model of A as an individual, I have to bring together all the cognitive models of A that I have in my calculus, as in the example set out in Table 6.1.

The emergence of relationship modelling in the primate clade does not require a new neural architecture. Social calculus is, essentially, a networked database with modelled individuals as nodes and relationships as links; and we already know that vertebrate brains work as neural networks (O’Connell and Hofmann 2012). This means that the Homo clade already had the cognitive architecture to hold social calculus as a network of nodes and links, rather than as lists. This makes adding or deleting nodes, and establishing, redefining and removing links, easy to do without disturbing the rest of the network. However, maintaining
A large neural network of relationships would be costly in terms of cognition, so we cannot assume that it was equally available to all members of the *Homo* clade, or that it is equally available to all modern human individuals. There may even have been minimum levels of size and complexity that brains had to reach before they became able to handle social calculus.

Social calculus works both ways: it allows me to understand the relationships between individuals, and it allows me to understand an individual through their relationships with others and with me. However, the more I need to understand A as the container of a complex set of relationships, the more I need a shortcut for the concept ‘A’ (a nickname), and the more I need to group others into relational hierarchies around A. I can shortcut my list of A’s relationships with others by grouping those others in terms of their holistic relationship with A; and I can then modify my personal list of single-argument Relationship-A relationships with others by grouping them in terms of my relationship with A and A’s relationship with them. To express this linguistically, I can represent A as a third-person cipher (*he*, *she*, *it* or *they* singular), and a group of individuals sharing a relationship with A becomes a third-person aggregate (*they* plural). Figure 6.1 shows how this could be represented. This process can be repeated for every person with whom I have a single-argument Relationship-A relationship, making the third-person cipher and aggregate representations into reusable non-specific placeholder terms.

Because this process is cognitive and not necessarily communicative, it can precede communicated language; and there does seem to be a precursor for this kind of social calculus in the socio-cognitive modelling of the chacma baboons we met in Chapter 2 (Cheney and Seyfarth 2007).

Table 6.1 List of my social calculus relationships between A and others

| A-Relationship(good)-B |
| A-Relationship(neutral)-C |
| A-Relationship(poor)-D |
| A-Relationship(good)-E |
| A-Relationship(poor)-F |
| A-Relationship(good)-G |
| A-Relationship(neutral)-H |
| A-Relationship(good)-I |
| A-Relationship(poor)-J |
| A-Relationship(neutral)-K |
These baboons live in a stable social hierarchy of individuals within families, giving deference to those above them to avoid confrontation, and expecting deference from those below them. The hierarchy is linear, because hierarchy between families takes precedence over hierarchy between individuals; and, by itself, it requires only simple Relationship-A modelling. However, baboons also learn from interactions between others in their group. They can remotely identify callers from their calls, and they pay more attention when, for instance, a threat bark from a subordinate is followed by a fear bark from a dominant. The hierarchy of families overlaying the individual hierarchy is also significant: after a confrontation, reconciliation with another member of the antagonist's family counts as a reconciliation with the antagonist. It seems, therefore, that chacma baboons cognitively maintain a network of relationships between others, albeit filtered through Relationship-A modelling. However, there is no indication that they communicate this network to each other.

The second reason why third-person pronominals are likely to have preceded other pronominals is that this form of reference is not based on the roles of speaker and listener. The first and second person, when used communicatively in a discourse, are in constant flux, depending on who is speaking (because my me is your you and vice versa). But a third-person reference can remain constant throughout a discourse, regardless.
of who is speaking: my Alf is the same Alf as your Alf, so my he is the
same as your he while it refers to Alf. Third-person pronouns are meta-
referential: they refer to a name or label that has been used previously
in the discourse, and that name or label in turn refers to a real person or
thing. So, once the name or label has been linked to a pronoun in a con-
versation, that link remains valid until replaced by a new link.

This also applies to cognition: once a mental place-marker has
been created to represent an individual, it can continue to be used in that
role. In its basic form, this is nominalisation, as each individual is likely
to be represented by a unique label. But in the social calculus form A-
Relationship-B, both A and B are meta-referential: they are placeholders
that can be filled by any nominalised label, which in turn refers to a cog-
nitive model of a real person or thing. This, once again, indicates that
the cognitive modelling of third-person pronominals can precede its
communicative usage. This is impossible for first- and second-person
pronominals, because they are markers of the communicative act; they
identify who is talking and who is being talked to, and have no cognitive
meaning outside of communication.

The third reason why the third person is likely to have been the
first instance of pronominalisation is that it is object-referring. While it
is called ‘third person’, it is also ‘first thing’: where I and you are neces-
sarily communicative, and therefore have an animate and active role in
the communication, the third person is only a thing referred to – and, in
the third person, people can be treated as inanimate things. The capacity
to blur the distinction between animate and inanimate is enshrined in
different ways in different languages. While languages like English pre-
serve some form of animate/inanimate distinction (pronouns he, she, it),
other languages like French do not (pronouns he, she). In the Basque lan-
guage, Euskera, in contrast, the proximity of the object or person is para-
mount (pronouns are represented by it [close], it [far] and it [between
near and far]); like French, there is no animate/inanimate distinction,
but where French treats everything as animate, Euskera treats everyone
and everything the same.

In cognition, distinguishing between animate, intentional beings
and inanimate, passive objects would seem to be a useful capacity.
Intentions, however, are only identifiable in terms of outcomes; and
‘passive’ is an outcome that can be attributed to both animate and inani-
mate objects. Additionally, inanimate objects do not even need to be per-
sistent, they can be transient events with outcomes; and the nature of
those outcomes can even challenge their inanimacy. A lightning flash
is just an event, but it can fell trees and kill people, which indicates an
apparent intention behind the event, and a possible being behind the intention. Identifying intention as a marker of animacy, and effect as a marker of intention, leads us to fallaciously reverse-link effect with intentionality, and intentionality with animacy. Third-person pronominalisation, like many cognitive capacities, comes with costs as well as benefits.

_They_ is, therefore, a cognitive event, the origin of which may precede the _Homo_ lineage, and which may be ancient in the primate clade. Its role in coordinating sets of social calculus pairings makes it cognitively valuable without communication: it contains all the features required for communicative pronominalisation without needing to be communicated – unlike the first- and second-person forms. It seems that, in terms of evolution, we may have long mislabelled our terms here: the first-person form to be used by humans was actually the third-person form.

**The origin of you and me**

Social calculus uses third-person pronominalisation to provide meta-reference without any need to communicate it. In contrast, first- and second-person pronominalisation do not have any function before we begin communicating our social calculus (Edwardes 2014). They are not a basic feature of social calculus, but they do simplify the exchange of social calculus models that include the receiver, _you_, or the sender, _me_. They also allow the sender and receiver to recognise the privileged nature of two particular aspects of social calculus: the _they_ that is the current receiver of my signal, and the _they_ that is the receiver of your signal.

For _they_, the cognitive concept was needed before, and therefore generated before, the communicative representation; but the communicative representation of _you_ was needed before its cognitive concept became useful or necessary. Yet it is impossible for a communicative representation to exist without its cognitive concept; so what pre-existing cognitive concept was available to be pressed into service as the cognitive concept _you_? The only clear candidate seems to be the cognitive concept of _they_: _you_ is a special case of _they_, a meta-meta-reference derived from the meta-referential value of _they_. The extra meta-level allows the internal short-form reference to an individual to include the communicative role of that individual; _they_ is a cognitive reference to a named other, or group of others; and _you_ is a cognitive reference to a named other, or
group of others, which is the object of a communicative act. You is just a privileged form of they.

So when, after we began communicating our social calculus, did you appear as a communicable representation? If we see you as a privileged form of they, then the answer has to be: quite soon after social calculus began to be shared, and very soon after the first sharing of an A-Relationship-B construct in which the receiver of the construct is either A or B. When sharing information about third-person others, the relationship construct is not intimate to either the sender or receiver; it is information about things happening ‘out there’. But a relationship construct that includes the receiver is only ‘out there’ for the sender – for the receiver, it is intimate and therefore privileged.

This leaves the problem of me, which is at the heart of this book. Just as you is a privileged form of they, so too is me: it is a meta-meta-reference derived from the meta-referential value of they. In this case, however, the extra meta-level allows the communicative short-form reference to be recognised as an internal cognitive representation of the individual themself; unlike you, what is being recognised is not a pre-existing element in the individual’s social calculus. When social calculus is just cognitive, the individual has no need for a model of me; the self is the unchanging and undefined centre of the calculus, the base on which the calculus is built. It is only when I am presented with someone else’s model of me that I need to acknowledge that there is a me to be modelled. This means that my cognitive model of me is not a product of perspicacious introspection, it is an amalgam of other people’s models of my self, a third-person representation of third-person representations which should not represent me in any sensible way, but which somehow do.

As the appearance of you and me must have been nearly simultaneous, the question of which came first is difficult to answer. The order of events works either way, because the shared you needs to be recognised by the receiver as me, and the shared me needs to be recognised by the receiver as you. Whichever came first, the appearance of the second from the first would have happened very quickly, and quite soon after social calculus constructs were being exchanged. It is possible (and probably best) to treat the exchange of social calculus constructs and the beginnings of you and me as simultaneous. However, the cognitive consequences of the concepts of you and me are interrelated but different; and the effects, particularly in the case of me, are wide-ranging in defining our species, our self and our selfness.
As we saw in Chapter 3, human cognition distinguishes between two types of human knowledge: etic facts, which are definitionally true but of which we are not necessarily consciously aware; and emic facts, which are true because we agree they are true (Brown 2004). Shared social calculus models are, themselves, emic facts, and their communication generates the emic facts of you and me. However, the emic fact of me relies on an internalised representation of someone else’s model of me; it is not me in any self-aware way, it is my awareness of your awareness of my selfhood. This makes it my third-person representation of my self, just another they model I can put into future what-if scenarios – including scenarios where my physical self does not survive. Yet the capacity to model my self into the future, and the capacity to model a future in which my model of my self no longer exists, somehow mitigates the very natural and visceral dislike of self-extinction. The ability to project my emic me into the future beyond my death is a very strange thing to be able to do. How can I cognitively visualise a world where I have no cognitive existence? And what is the evolutionary advantage of being able to do so? Recognition of the emic me may actually be a two-edged sword for the individual, in that the capacity for social self-modelling damps down our primate Machiavellian self-interest.

Self-modelling works backward in time, too: the emic fact of me allows me to relate my current model of me to my memories of me, creating the impression of a continuous self that exists through time. I can also receive modelled memories and other stories from other people about times and places in which I have never existed; and I can relate my current model of me to those stories. I can be with Harold at Hastings as he takes one in the eye (an event that probably never happened, and certainly not in my experience); I can be on the steps with Sidney Carton as he does a far, far better thing than he has ever done; I can even be with our early ancestors as they evolved into modern humans. The emic me is more than just a placeholder for a cognitive model of the self, but it is less than recognition of the self itself: more than I think, but less than I am.

This dual deception of the emic me (that the modelled self explains the self, and that the modelled self is the self) gives us a more nuanced understanding of self-sacrifice. As we saw in Chapter 5, there is a tendency to dispassionately see the self in our models of self-sacrifice as synonymous with the self that is sacrificed; but this is not the case. It is easy to sacrifice the emic me because it ‘lives on’ (or at least has purported cognitive existence) after death, and ‘existed’ before birth; it has a timeless quality that death cannot affect. In this cognitive environment, self-sacrifice becomes more than acceptable; it can even be a sought-after,
or valued, quality. We do not need to go as far as Martin Luther King Jr when he said, ‘if a man has not discovered something that he will die for, he isn’t fit to live’, but we probably do recognise a personal mental line beyond which death is preferable to living. Yet we also do not recognise the crossing of this mental line in others, treating suicide (especially in the West) as a foolish or bad choice (which, in evolutionary terms, but not in self-modelling terms, it is).

Here we are faced with a conflict between the two different emic me’s, the social self-model and the cultural self-model: for the Social self, the emic me, it is the costs and benefits to the is self-model that need to be assessed; for the Cultural self, it is the costs and benefits to the should be self-model that matter. These can be very different things; and there is always the etic subliminal evolutionary drive toward self-survival to be taken into account, too. In self-sacrifice, the social self-model and cultural self-model seem to be working together against the evolutionary drive; in suicide, the social self-model is, by itself, against the cultural self-model and the evolutionary drive (at least, that is the case in cultures where suicide is morally frowned-upon). However, while the outcome of this three-way conflict is not easy to predict, the social self-model wins enough times to make suicide a recognisably human phenomenon.

**The origin of possession and the possessive**

Pronominalisation in many languages also gives us the power to extend our umbrella of selfhood to include non-self objects: things that are beyond the limits of my self can nonetheless become my things. The way this is done varies between languages; there are many different ways of possessing in the world’s languages, both in terms of what is possessed and how the possession is expressed.

For instance, several languages differentiate between alienable objects (things that can be possessed by others, like *John’s dog*) and inalienable objects (things that can be possessed only by the individual, like *John’s dream*). Some languages differentiate between inherent possession (parts of a whole, like *John’s finger*) and non-inherent possession (where the two things remain equal and separate objects, like *John’s dog*). Some languages treat some things as unpossessable, and some treat the possession of places as different from other forms of possession by using a special locative possessive.

In English, we do not officially differentiate between alienable and non-alienable, or inherent and non-inherent, but there is nonetheless
some differentiation. We can use the *of* form of possession with inalienable but not alienable possessions: *the dreams of John* feels acceptable (although it has two very different meanings), but *the dog of John* does not (although *that dog of John’s* is acceptable). There is also some differentiation between non-inherent and inherent possession: *John’s dog* requires the possessive form (*’s*, as in *Alf’s nose*), but *the family’s dog* can be reduced to *the family dog*, possibly because the dog is an inherent part of the family in a wider sense. This double-noun formation is very common in English (for example, *history test, business contract, fly spray, bicycle wheel*), and it is unlikely that inherent possession can be invoked to explain all double-noun forms; but it does help with some.

English does not seem to have any restriction on what can be possessed, treating even the ultimate supernatural entity as a personal possession (‘My God, my God, why have you forsaken me?’ – Psalm 22:1, *Holy Bible, New International Version*). However, we do have an example of a special locative possession form: the ellipsis in *I’m going to John’s* indicates a place possessed in some way by John, such as ‘house’ or ‘home’. The way possession is marked in different languages is also variable. English uses: prepositions (*of* and *for* particularly, but not exclusively); a possessive form (*’s*); double nouns; or noun phrases (such as *the dog owned by John* or *the dog John has*). Japanese uses a participle *no* following a noun phrase to indicate the noun is a possessor (*Jon no inu*, John’s dog, for instance); Latin uses a genitive case-ending on a noun to indicate it is a possessor (for example, *Pax Romanorum*, the Roman peace); and languages in the South American Cariban group use a case-form to indicate the possessed thing rather than the possessing thing. To date, however, no language has been identified that does not indicate possession in at least one way.

Possession may be ubiquitous because it is just a particular form of A-Relationship-B social calculus: it is an unequal relationship between a person and a thing, where the person has a dominance relationship over that thing. It possibly comes directly from a social calculus construct that marks a dominance–submission relationship between two people, but replaces the submissive person with a thing. This possessive extension of the A-Relationship-B form is one of many ways the form can be elaborated: it can express other relationships between things and people, or relationships between things and other things. The relationships themselves can also be replaced by actions (existential actions of being, material actions of affecting, relational actions of representing, behavioural actions of performing, verbal actions of describing, and mental actions of ideating) or other links (cultural links like marriage, economic
The expression of possessive relationships is one of many ways in which the simple sentential communicated form of A-Relationship-B can be pressed into new uses: out of this A-Relationship-B social calculus construct can come all the simple grammar we use in modern language.

One common way of expressing possession is to use a pronominal to replace a noun phrase (mine / ours / yours / his / hers / theirs). Just as the basic pronominals all trace back to the cognitive meta-referent of they, so the possessive pronominals trace back to the use of they in an A-Relationship-B construct, where the relationship is one of possession. Social calculus involves sets of paired relationships between known individuals, and the use of they allows individuals and groups to be represented by a generic placeholder; possession is just another type of paired relationship, and the same meta-referent pronominal properties apply. So just as the three persons in a communicative act can be represented singularly or plurally by pronominals, so possession can be expressed in all three persons in singular and plural forms.

Possession seems likely to be one of the earliest ways in which the exchange of social calculus was broadened to include other communicable information; and the famous Blombos crayon (Henshilwood et al. 2002) looks like an early exemplar of possessive messages – perhaps the earliest known. A simple cross-hatching on a piece of soft ochre, probably used as a crayon to create red pigment, indicates a relationship between the crayon and at least one human. We cannot know what the cross-hatching means, but we can say with confidence that it is an indicator that, about 75,000 years ago, this particular piece of stone was important enough to be personalised for identification.

Possession is, however, more than just an extension of the human way of seeing the world; its role in communication changes the world. Where first-person possession is expressed (for example, ‘my rock’), it poses a challenge to the worldview of others. It is an assertion about reality, but it is only as real as others allow it to be. Possession plunges us into Karl Popper’s Three Worlds (1967, Chapter 4), which constructs the human view of knowledge on three levels: what continues to exist even in the absence of humans, such as rock, constitutes World 1 (actuality); that which exists only inside human heads, such as my, is World 2 (virtuality); and that which has actual existence without humans but has meaning only because of humans, such as crayon, is World 3 (reality).

It is difficult to represent these three worlds using words, because every word is a negotiation toward meaning – it exists in World 2 and can only
represent Worlds 1 and 3, not exist in them. The word *rock* represents an actual rock without being one, and the word *crayon* represents the idea that a particular rock can be described by role as well as by substance. The word *my* is an assertion of possession in World 2, but it represents different things in Worlds 1 and 3: in World 1 it represents actual, physical possession, and in World 3 it represents agreed, or legal, possession. These three meanings together form our understanding of what constitutes ‘good’ possession: it is asserted, actual and agreed. Without actual possession, it is merely desire, and without agreement it is usurpation; but if all three are present, then our World 2 virtual concept of possession has changed our World 3 real model of how World 1, the actual world, works and is ordered. Language may not have changed the actual world, but it has changed our concept of the world, and therefore our relationship to it.

**The origin of recursion and reflexivity**

In Chapter 3, we looked very briefly at the topic of recursion in terms of the sharing of the A-Relationship-B constructs we have been offered by others. By attributing the whole construct to the original author, we are able to pass on what may be deceptive information without risking our own reputation. This construct, ‘C said A-Relationship-B’ is hierarchical in that the basic construct is contained within a larger construct, [[[A-Relationship-B] by-C]]; and it is recursive in that it can be further nested as [[[[A-Relationship-B] by-C] by-D]] by D. In English, this makes a sentence like ‘Del said that Gemma told him that Alf likes Beth’. In theory, the *by-x* form can be added into the construct an unlimited number of times, but in practice human minds can usually only handle five or six levels of nesting (Dunbar 2004, Chapter 3). The *by-x* form is not the only type of recursion possible in language; but, like self-modelling and first- and second-person pronominalisation, it is probably an early one.

Why should recursion be so significant in terms of language origins? The answer lies in the recent history of linguistic research. For 60 years, linguistics has been strongly influenced by the theories of one man, Noam Chomsky. His theoretical viewpoint, that language is a specialised cognitive computational system (1995b), has been extremely influential and has affected linguistic theory profoundly. It has also encouraged three generations of researchers to adopt four counterintuitive views about language: first, that language is for thought, not communication; second, that language cognition is just for language – it is a specialised system
that is not used for other types of thought; third, that we must, if we are to communicate using language, all have the same language engine (Universal Grammar) in our heads; and fourth, that the language engine is exclusively human – no other animals have anything like it (Chomsky 2007). This position has become known as generativism, from the idea that the engine, Universal Grammar, generates all our language capacities and all human languages.

The literature in support of the generativist position is impressive (Chomsky himself has written, or been involved in writing, over 60 linguistics books). But evidence in the last 20 years from psychology (such as Adornetti and Ferretti 2014), neurology (Arbib 2005), physiology (Evans and Levinson 2009), biology (Bickerton 2014), archaeology (Hoffmann et al. 2018), linguistics (Beckner et al. 2009), animal studies (Sergiel and Ferretti 2005), child studies (Ibbotson and Tomasello 2009), evolutionary theory (Wacewicz 2016) and complexity theory (Kirby et al. 2008) – even economics (Alonso-Cortés 2006, Chapter 4) – is not providing the necessary support for the idea of a genetically specialised and species-exclusive language engine.

The situation is complicated by the fact that not all generativists believe the same things about language, and Chomsky himself has altered his position several times. Most drastically, in 1995 (1995a) he abandoned his previous theoretical suite of Principles & Parameters and Government & Binding (for example, Chomsky 1982) to concentrate on a reduced package involving just two principles: MOVE and MERGE. He has now largely dropped MOVE as a separate principle, concentrating on the single feature of MERGE as the cognitive capacity that separates humans from the rest of nature, and language from other communication systems (Hauser et al. 2002). Another name for MERGE is recursion.

For Chomsky, MERGE is a human-only capacity, the result of a mutation in a single individual sometime between 200,000 and 60,000 years ago (Berwick and Chomsky 2016, Chapter 3). This mutation revolutionised the individual’s life, giving them a reproductive advantage that passed down through the generations, out-competing unmutated versions of the affected gene because the cognitive capacity of MERGE provided so many fitness advantages. How a cognitive capacity creates a reproductive advantage has not been fully explored; and how the gene managed to replace the entire unmutated stock of genes has not been examined in a systematic way. Even the wide date range raises some questions. We now know that the single migration out of Africa by Homo sapiens 60,000 years ago was probably preceded
by an earlier migration about 120,000 years ago (Bae et al. 2017), a
date supported by recent discoveries about *Homo sapiens* in China (Liu
et al. 2015) and Australia (Clarkson et al. 2017). We also now know
that modern *Homo sapiens* was present in Morocco 300,000 years
ago (Hublin et al. 2017) and had spread to Omo Kibish in Ethiopia by
195,000 years ago (Sisk and Shea 2008), a distance of over 5,000 km.
So, if the new mutation had occurred after 200,000 years ago, it
would have needed to spread horizontally into existing populations
by interbreeding, as well as vertically down the generations. It would
also have needed to coexist with the genetic heritage of other species
of humans: we now know that there were signs of interbreeding
between *Homo sapiens* and both *Homo neanderthalensis* (Green et al.
2010) and *Homo denisova* (Reich et al. 2010). In Europe, on average,
our genome is up to 4 per cent Neanderthal; and in Asia, the human
genome includes about 0.2 per cent of Denisovan genetic material as
well as the Neanderthal content. African humans have no Neanderthal
or Denisovan genetic material, while the Melanesian genome is up to
8 per cent Denisovan and Neanderthal combined (Bustamante and
Henn 2010).

This leaves a problem for the generative model of language
origins. If Chomsky is right that MERGE is the result of a mutation, then
he must be wrong about the timing: the mutation must have happened
early enough in the history of *Homo sapiens* to ensure it reached
Australia 65,000 years ago, because interbreeding between outsiders
and Australian Aboriginal humans was slight-to-non-existent for all
but the most recent 300 years. It would also be a better explanation if
the mutation happened before 200,000 years ago, to give it a chance
to spread through the whole *Homo sapiens* population before *Homo
sapiens* started spreading through the world. If, however, the timing
is adjusted, then there is the problem of how this mutation changed
the fitness profile of the species. If it had a major effect on fitness (as
the generativists claim for cognitive recursion), then it should spread
quickly, and evidence of what that fitness change did to humans
should be evident in the archaeological record; but the technological
and social record from archaeology gives no clue as to any cognitive
changes before about 100,000 years ago. If it had only a minor effect,
then why should it spread quickly through the population? The success
of any mutation is measured in terms of relative numbers of offspring,
so if it is to spread quickly both vertically through selective inheritance,
and horizontally through interbreeding, it has to be quite remarkable
in its effects.
There are three ways out of this problem. The first is to treat MERGE (recursion) as non-genetic: it could be a cognitive trick that some humans mastered and then shared; this trick could spread culturally – and, therefore, much more quickly than if it were a mutation. This, however, still leaves the questions of why and how this cognitive trick was first recognised, why it proved so useful, and how it was shared without recursive language already existing. The second way out of the problem is to treat MERGE as an ancient genetic capacity that happened to be communicatively useful when language appeared, but which was available and useful in cognition long before language appeared. However, the question here is: what purpose did recursion in cognition serve? No cognitive cost could survive for long without a countervailing fitness benefit. The third solution is to treat MERGE as a communicative potential already cognitively present in hierarchical social calculus, a potential that was realised only when that calculus was shared. This is the solution offered here: the ability to cognitively model individuals within groups while separately identifying the individual and the group is potentially hierarchical; so when the communication of social calculus actually begins, the recursive tools are already in place. While shared social calculus models are initially linear, they soon require the ability to share one person’s model of another person’s model of a relationship, indicating the ownership of the information in the model at each level. This sharing of information sources is hierarchical and, within a limited definition of the term, recursive.

Robin Dunbar describes these nested tagged models as ‘a hierarchically organised series of belief-states’ (2004, 45); and he goes on to show that, while the capacity to nest individuals and groups within groups seems to extend to at least seven levels without diminution in understanding, the nesting of relationship models becomes ineffective and highly error-prone beyond five levels of nesting. Recursion is not, as Wilhelm von Humboldt (1836 [1999], 91) says (and Chomsky is fond of quoting), ‘infinite employment of finite means’, it is constrained by the capacities of the human brain, which are not as amazing or unusual as we often pretend. Caballero et al. (2018) raise an important issue about infinitely recursive iteration: cognition is about decision points. At some time in the process, the formulation of an idea must result in a formed idea; and this cannot happen without a stop-marker on the iterative recursion formulating the idea. Decision-making does seem to rely on iterative processing, and a large number of those iterative processes are recursive; but that does not make them infinite, or even possibly infinite.
One aspect of recursion that is not often explored using MERGE is reflexivity, the capacity to refer an activity back on itself so that the instigator of the action is also the receiver. In the notation used here, this would be an $A_1$-Relationship-$A_2$ construct, a recognition that a they has the same ability to model themself as I do. To express reflexivity in English pronouns, we use possessive+self ($myself$, $ourself$, $ourselves$, $yourself$, $yourselves$) or, in the third-person, object+self ($herself$, $himself$, $itself$, $themself$, $themselves$). This differentiation of form is neither explicable nor fixed – some dialects allow $meself$ or $hisself$ or $themselves$, but never $weselves$, $youself$, $yourselves$ or $sheself$. They do all, however, reflect the fact that $A_2$ is a model of $A_1$ made by $A_1$ – which is itself a model. This is clearer in the semantically complex utterance {name 1}- {has a good relationship with} - {name 1} ($John$ loves $John$). This reflexivity is used to indicate disapproval of the relationship, or surprise that it is possible for anyone to love $John$; but it is done by intimating that $John$ actually loves $John$'s model of himself, and this is a wrong thing for $John$ to do. In practice, of course, when I utter $John$ loves $John$, I am really saying ‘my model of $John$ loves my model of $John$’s model of $John$. I leave you to work out what is happening in ‘$John$ loves that he loves himself’; but the fact that you are able to extract meaningful information from these six words about my intentions, $John$’s intentions and my relationship with $John$ shows how the sharing of social models, accompanied by recursion, creates a powerful information machine.

Self out of language, language out of self?

This chapter has approached selfhood as a communicative issue, and particularly as a language event. If the story of self this book proposes is close to what actually happened in our prehistory, then the human capacity for language is heavily implicated; and an understanding of what language is becomes necessary. However, we are immediately faced with a definitional problem: what type of communication counts as language? This question is not as easy to answer as it should be. For some linguists, language is how only humans communicate; so, by definition, it is a human-only capacity. For generativists, the key ingredient is recursion; so any human communication system that cannot be used recursively does not count as language. For other linguists, language happens whenever communication is sufficiently complex; so any communicative act could count as language if it is complex enough, whether it is performed by humans or by another species.
In this chapter, the term ‘language’ has been used cavalierly, with no particular definition attached. This is deliberate, because what counts as language is something of a red herring in the search for selfhood. To begin exchanging social calculus models, there must have been a pre-existing communication system that allowed meaningfully differentiated segments of a signal to be brought together to make a meta-signal. Fortunately, recent evidence shows that this might be an ancient capacity: marmosets, which are in a lineage with which we shared a common ancestor 40 million years ago, appear capable of vocalisations that ‘do not consist of one discrete call pattern but are built of many sequentially uttered units, like human speech’ (Pomberger et al. 2018). While marmosets are unlikely to have a capacity for social calculus, a call pattern built of sequential units is all that is needed to share social calculus models. Whether we class the exchange of social calculus models as sufficiently linguistic to be called language is immaterial to the fact that all the cognitive resources we needed to share social calculus models were probably already present before Homo sapiens appeared.

However, a knowledge of selfhood is not needed to communicate social calculus models. Indeed, if a capacity to model the self as a third person, a dispassionate self, allows in the capacity for self-sacrifice, it can be argued that self-modelling, by itself, makes an individual less fit than an individual who does not self-model. Only if self-modelling is a secondary outcome of something that increases fitness should it become common in a species. Could sharing social models enhance individual fitness sufficiently to counteract the negative effect of self-sacrifice? The answer probably lies in the increased levels of trust that the sharing of models engenders, and the enhanced knowledge of how the individual’s group works; but that is not an issue that needs to be pursued here, for we have discovered a viable path from self-serving, self-free Machiavellianism to selfless, self-modelling selfhood.

The journey on from selfhood is of interest in terms of language itself. Sharing second-hand models requires recursion to enter the equation, and this allows human communication to become increasingly complex. This is significant for generativists, because it crosses the Rubicon they believe exists between pre-language communication, or proto-language, and full language. However, whether what happened before communicative recursion counts as language or not is a matter of opinion. We can define language in whichever way we wish, to fit our theory of how humans communicate; it makes no difference to the nature of the communicative act itself. For selfhood, we needed an effective,
segmented and differentiated communication system already in place; and we needed a cognitive capacity for social calculus. From selfhood, we gained a complex and iterative communication system that any linguist would be happy to label language.

Notes
1. The terms Actuality, Reality and Virtuality are my own terms for Popper’s Three Worlds.
2. What von Humboldt actually said was: ‘Sie muss daher von endlichen Mitteln einen unendlichen Gebrauch machen’. So a more literal translation would be: ‘it must therefore from limited resources generate unlimited usage’. This is sufficiently different from ‘infinite employment of finite means’ to argue that von Humboldt and Chomsky may not be talking about the same thing.