

---

## I Am What I Am: the origins of grammar in self-modelling

---

Language is often identified as a defining difference between humans and nonhumans; but is it a feature with its own evolutionary explanation, or is it an adjunct of some other trait, evolved for reasons only indirectly related to language? Despite Noam Chomsky's view that language evolved for thinking<sup>1</sup>, it is generally believed that language is a product of socialisation, the need for individuals to exchange social information. And, if we look at the social information exchanged, we see forms of modelling that are absent from signalling systems used by other animals.

First, there is modelling of the relationships between others: the simple sentence Alf likes Beth requires the receiver not just to recognise a single referent or context, but to bring two conceptualised individuals together and establish a relationship between them. Second, there is modelling of events in the past and future: where other animal signalling is dominated by current events, these play only a small part in human discourse. Third, there is modelling of what-if: we can discuss not just future probabilities and past certainties, but events that are unlikely to, or did not, happen. Fourth, there is modelling of the intentions of others; and fifth, there is modelling of our own self.

The question of selfhood has long been recognised as vital to understand what makes us human. Even before Darwin's theory of evolution by descent, Wilhelm von Humboldt stated that:

... the customary demands of mankind are satisfactorily met by forces of nature and by the mechanical continuation of human activity. But the appearance of a greater individuality in individuals and in peoples, practically inexplicable by any derivation, interferes suddenly and without warning with the course more obviously determined by cause and effect.<sup>2</sup>

As humans, we are each aware of our own individuality. However, I am aware not just of myself, not just that you have a model of myself, but that others can have models of the relationships between me and you; and, most importantly, that the viewpoint of those others can be adopted by both me and you. I must simultaneously have an awareness of my own self-image, the image of me maintained by you and others, and the images of the relationship between me and you held from a possible fourth-person viewpoint. *Homo sapiens* seems to be the only species that does this.

### Model Making

Other species can certainly make models. In particular, chimpanzees and bonobos are, like us, Machiavellian thinkers. We share with our closest relatives the understanding that the actions of others can be influenced, that others are useful or dangerous, and that others have relationships with each other. To express this as a calculus, we have emotive mental constructs of our relationships with others which range from fear

---

<sup>1</sup> Noam Chomsky, *New Horizons in the Study of Language and Mind*, pp27-29

<sup>2</sup> Wilhelm von Humboldt, *Linguistic Variability and Intellectual Development*, p4

through to affection; and emotive mental constructs for all other relationships in our group.

We use these relationships to predict likely behaviours which can in turn modify our own behaviour: I fear Alf, but I have a very good relationship with Beth, and Beth has a good relationship with Alf. If I stay close to Beth, Alf is unlikely to attack me because Beth is more likely to support me than Alf, and Alf therefore risks his good relationship with Beth. And, most importantly, I can assume that Alf knows this, too.

What is happening here? We can express it in human terms as the ability to make models. I am able to model the relationship between Alf, Beth and me in terms of the separate relationships between Alf and Beth, Alf and me, and Beth and me. Two types of knowledge are needed: knowledge of how you (the immediate other) react to me, and knowledge of how others react to you. These both involve second order intentionality – the first is my thoughts about your thoughts, and the second is your thoughts about their thoughts<sup>3</sup> – but they are very different types of knowledge. In the first, your thoughts and intentions directly affect me; in the second, they only indirectly affect me through my own intentions towards you and towards the other.

There are also syntax or calculus considerations: my relationships with the world rely on a constant, *me*, relating to variables out there. In contrast, relationships between variables out there have no fixed constants. To understand relationships which do not involve *me* directly, I need to understand two types of second-order intentionality: I need to know about your intentions to others as well as my intentions to you.

We can see in these two model-making abilities the two universals of language: my relationship with others involves a simple cognitive correlation between differentiated objects (other individuals) and emotive states. This gives us the action-object distinction. The reaction of one individual to another is a propositional relation between two objects, which gives us the subject-verb-object propositional form. This has implications for a Formalist Linguistics position: yes, there are universals in language, but they are universals inherited from cognitive structures likely to have been available to the common ancestor of *Pan* and *Homo*. As James Hurford points out, the predicate-argument (object-action) distinction is neurologically based, and it is a capacity available to apes and monkeys as well as humans<sup>4</sup>.

## The Self and Language

When reviewing the issue of self in language the first question that needs to be addressed is simply, what is the self? Or, to put it in more immediate terms, what is me? This is not the same question as the ubiquitous “who am I?”, which aims to identify an intimate self, mostly through introspection. “What is me?” attempts to describe the self as an externalised model – the viewpoint is not that of the interested self but of a

---

<sup>3</sup> Robin I M Dunbar, *The Human Story: a new history of mankind's evolution*, pp45-51

<sup>4</sup> James R Hurford, The Neural Basis of Predicate-Argument Structure. In *Behavioural and Brain Sciences* 2003 26 (3), pp283-284

disinterested fourth-person. Julian Jaynes describes the modelled self as the metaphor ‘me’, in contrast to the analog ‘I’<sup>5</sup>.

Susan Greenfield answers the question “what is me?” by reference to an always-present self and an awareness of that self through linguistic self-reference:

... language gives us a symbol for something that normally does not make inroads into our senses, simply because it is always there: one’s self. As soon as we have a simple word for ourselves then we can inter-relate the self in context. We can become self-conscious. This self-consciousness, combined with the ability to escape from the here and now, is surely what really distinguishes us from almost all other animals, as well as seemingly inhuman human infants.<sup>6</sup>

In Greenfield’s formulation the self in context (the metaphor ‘me’) is a conscious reflection, or model, of one’s self (the analog ‘I’). We are able to describe our model self because it is a direct product of our conscious cognition; but we can only imperfectly describe our true self – others can describe our true selves more accurately than we can. Tomasello sees this process the other way around: it is our increasing self awareness in childhood that creates our knowledge of the capacities for others<sup>7</sup>; but this is not a common viewpoint. David Dunning, Chip Heath and Jerry Suls show that self-judgements on intelligence have a low correlation with real intelligence, completion of our tasks is always behind our estimated completion, and we are incurable optimists when it comes to our own health<sup>8</sup>. Roy Baumeister, Jennifer Campbell, Joachim Krueger and Kathleen Vohs show that our self-esteem does not match well with the esteem given by others, and it also does not correlate well with academic achievement<sup>9</sup>. Benjamin Franklin said “there are three things extremely hard: steel, a diamond, and to know one’s self”. This would appear to be a reasonable view in light of the evidence.

It seems likely, therefore, that we understand others better than we understand ourselves, and most knowledge of ourselves comes not from introspection but from modelling the minds of others and their intentions towards us. Self awareness is better served by comparison than by introspection.

The attempt to “know thyself”<sup>10</sup> is a conscious cognitive act for humans, it is not the same as the self interest, or selfishness, that Richard Dawkins sees as underpinning the evolutionary process. Dawkins’ selfishness is not concerned with activities at the phenotypic level but at the genetic level. If they are expressed at the phenotypic level then they are expressed as autonomic responses to environmental stimuli. Genes do not choose to be selfish, they are selfish because only selfish genes survive: genetically inspired actions that favour the phenotypic self and its reproduction lead directly to

---

<sup>5</sup> **Julian Jaynes**, *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, pp59-65

<sup>6</sup> **Susan Greenfield**, *Brain Story: unlocking our inner world of emotions, memories, ideas and desires*, p169

<sup>7</sup> **Michael Tomasello**, *The Cultural Origins of Human Cognition*, pp70-77

<sup>8</sup> **David Dunning, Chip Heath and Jerry M Suls**, Picture Imperfect. *Scientific American Mind* Vol 16 No 4, 2005, pp20-27

<sup>9</sup> **Roy F Baumeister, Jennifer D Campbell, Joachim I Krueger & Kathleen D Vohs**, Exploding the Self-esteem Myth. *Scientific American Mind* Vol 16 No 4, 2005, pp50-57

<sup>10</sup> Said to be written over the door of the temple of Apollo at Delphi

genotypic survival, actions that favour others do not. There may be an indirect path to survival by favouring others, but that indirection has to be advantageous enough to outweigh the direct path of non-co-operation before altruism should appear<sup>11</sup>. There is no intentionality in Dawkins' gene model of selfishness, any more than there is intentionality in a hurricane: both are natural phenomena and controlled by fully explicable rules external to the phenomenon.

Genetic selfishness is different to sense of self, which in turn is different to self awareness. Genetic selfishness is a default state which does not require consciousness. It ensures survival, and requires no more knowledge of the self than that the world is divided into self and not-self. What is inside the line, the self, is the ends; the rest, the other, is just means. Of course, with a binary model such as this only one of the items needs to be defined. It would seem that the self is the easiest to define, but it is also the least useful. The self is that part of the Universe that is already under control, so it is trivial; much more important is that part of the Universe that has to be manipulated and negotiated. A feature of genetic selfishness is therefore likely to be a lack of comprehension of the self. *Sense of others* gives immediate advantages, it allows an organism to subvert the survival of those others to its own purposes; *sense of self* gives no such immediate advantages.

Although sense of self does not have direct advantages, it does have indirect advantages: it allows an organism to exercise choice between strategies. In any situation there is usually more than one viable strategy of advantage to the organism, and the ability to choose effectively between them maximises the advantage of each strategy. Adopting a single strategy for a situation relies on there being no organism with effective choices at the other end of the strategy; as soon as a single strategy approach is met by a variable response (which an effective choice allows) then it ceases to work as effectively, and it can even become counter-productive. However, in order to make choices an organism must have a rudimentary understanding that there is a self to make the choices. This understanding does not have to be a conscious act, it need be no more than a recognition at the genetic level that the other half of the binary relationship, self and non-self, exists.

As soon as there is an understanding that there are choices, however, it becomes advantageous to model those choices onto other organisms. If I come from a lineage that has been successful because it is able to make choices, then my immediate rivals are also likely to come from that lineage. The ability to anticipate their choices is the next logical step in gaining a reproductive edge. It becomes advantageous to develop *other awareness*, a knowledge that others have choices that can affect my choices, and the ability to anticipate those choices. The model of the choices available to other organisms (and their possible responses) can be no greater than the knowledge I have of my own choices, so an increase in the number of my choices leads to more sophisticated models of others. These are two different types of knowledge, however: my choices are unconscious, they are selected by emotional bias and can be dictated by my feelings; in contrast, my analysis of your choices has to be, on some level, a conscious act – I have to be aware of your choices in order to cognitively evaluate them. We need conscious

---

<sup>11</sup> **Richard Dawkins**, *The Selfish Gene*, ch5

cognition to choose a course of action based on the possible actions of others, but we can rely on our autonomic responses to carry out that course of action.

In analysing your choices I am trying to evaluate your intentions: which choice will you make? But I am not evaluating my own intentions – they emerge from my feelings about your intentions and do not need conscious analysis. This means that the aboutness in this intentionality is not about me but about my model of you; there is no Theory of Mind in other awareness. To use Tomasello's definitions, I am not treating you as a mental agent; and I'm not treating myself as an agent at all. *Other awareness* allows me to generate an increasingly sophisticated map of your intentions, but it provides only a limited and difficult-to-expand toolbox to deal with those modelled intentions. It provides Machiavellian intelligence, but with no immediate way to allow that Machiavellian intelligence to become generative intelligence.

With humans and language we have a new type of self to be recognised. Michael Tomasello and Josep Call label this *self as social agent*<sup>12</sup>; Steven Pinker labels it *self-knowledge*, giving it as one of the three definers of consciousness (the other two being *sentience* and *access to information*)<sup>13</sup>; and Jerome Bruner calls it *the transactional self*<sup>14</sup>. Here, this final stage will be labelled *self awareness*, the fourth stage of modelling. Somehow humans are able to extrapolate from making Machiavellian models of others to making models of ourselves, which allows us to conceptualise ourselves as if we are looking in from the outside. The picture we have of ourselves is often inaccurate, but the ability to generate it at all is an evolutionary conundrum: how have we become able to take a disinterested viewpoint of ourselves?

Self awareness is a process that makes possible what we do with language. The reason why internalised modelling became externalised in language is a matter of human acculturation and will not be explored here, but self awareness also establishes the rules which make language so powerful: it allows models of the intentions of others to become recursive models of the intentions of the self and others. The self becomes identified with the other to such an extent that their roles in language are interchangeable: I can see myself simultaneously in three ways: as the instigator of an action, as the recipient of the action, and as part of the context in which the action happens – I can occupy the first, second and third person roles in an utterance. I can also see you as instigator, recipient or context, and both of us can be replaced in a construct by third parties: yesterday's "you" becomes today's "they", which I can tell another "you" about.

The dual self image permitted by this modelling is prevalent throughout language. It is even possible to identify the roles in specific English usages. For instance, the apparently interchangeable reflexive forms, *I hate me* and *I hate myself*, seem to have different roles in the identification of the nature of self<sup>15</sup>, with *me* representing Jaynes' analog "I" and *myself* representing the metaphor "me".

---

<sup>12</sup> **Michael Tomasello & Josep Call**, *Primate Cognition*, pp337-338

<sup>13</sup> **Steven Pinker**, *How the Mind Works*, pp134-136

<sup>14</sup> **Jerome Bruner**, *Actual Minds, Possible Worlds*, ch4

<sup>15</sup> **Martin Edwardes**, I like Both Myself and Me, *Proceeding of the Camling Conference 2003*

Self awareness would seem to be a function of socialisation. Kenan Malik argues that self awareness is intimately tied to language and social living – unless we have the knowledge that others have intentionality we can never have knowledge of our own intentionality:

Humans, however, are symbolic creatures, with language, self awareness and a social existence. These three phenomena are intimately interconnected. Language can only exist in a social form, but it also helps create the possibility of a social existence beyond simply the kinds of individual interactions that animals experience. The existence of a community of beings possessing language allows us to make sense of our inner world, and hence to become self-conscious. At the same time, I am only conscious of myself insofar as I am a member of such a community.<sup>16</sup>

This socialisation dimension means that two further capacities become possible with self awareness. The first of these is anticipation, or a second-guessing between intentions: I know your options, and I know my options, so I should choose the one that gives me the best result in response to your best choice. But then you know my options and you know your options, so you may choose the option that gives you the best result in anticipation of my best response to your best choice, so I should choose the best response to that option... There is a recursion between your intentionality and mine within both of our minds, and this recursion is one of the capacities that enable language – as the Hauser, Chomsky and Fitch model predicts<sup>17</sup> (although it is not the only condition necessary for language).

The second question (or series of questions) that self awareness makes possible is speculation on the intentions of others to each other, with no direct reference to the self's own intentions. Other individuals are modelled not just as animate agents with linking relationships, but as mental agents with their own intentions. Modelling the intentions of others is not done to identify strategies which are directly useful to me, but simply to identify what is going on. It is this capacity that enables and informs the insatiable and disinterested curiosity of humans<sup>18</sup>.

So out of self awareness come these three further types of cognition:

- **Reflexion:** the analog “I” is separable from, and interchangeable with, the metaphor “me”.
- **Anticipation:** there is a recursion between intentionalities.
- **Speculation:** the intentionality of others is modelled in language into both the recipient of the action and the instigator – roles become interchangeable.

The four stages of mental modelling take us from the reactive state of genetic replicators to the human ability to anticipate the thoughts of others. For Paul Bloom this last stage, self-awareness, is an important feature not just in being human but in language learning

---

<sup>16</sup> Kenan Malik, *Man, Beast and Zombie: what science can and cannot tell us about human nature*, p220

<sup>17</sup> Marc D Hauser, Noam Chomsky & W Tecumseh Fitch, The Faculty of Language: what is it, who has it, and how did it evolve? In *Science* vol 298 22 November 2002, pp1569-1579

<sup>18</sup> Ian Stewart & Jack Cohen, *Figments of Reality: the evolution of the curious mind*, pp163-164

itself: children do not learn words by a process of association, they learn them by inference from the intended meaning of others. This means that children, when they begin to utter their first associative words, already have sufficient modelling ability to understand that the word-sign is a negotiation between them and other people. They also understand enough about intentionality to know that the meaning of a word-sign is in the intention of the speaker (sender), and it is the role of the listener (receiver) to try to apprehend that meaning<sup>19</sup>.

Children are able to attain a Theory of Mind because they are born with a theory of theory. They seem to understand implicitly the process of thesis-antithesis-synthesis which is the heart of human scientific method. They apprehend the world, make models of it, check those models against new realities as they arise and modify their models appropriately. Alison Gopnik, Andrew Meltzoff and Patricia Kuhl call this “the scientist as child”, comparing the childhood modelling which builds adult competence with the human adult ability to continue modelling into adult life. Humans continue to play in the “mental sandpit” of modelling throughout their lives<sup>20</sup>.

Self modelling raises the issue of temporality: humans have an image of themselves as continuous with their past selves and future selves, but they are also able to see those past and future selves as if they were other people. This is something that non-linguistic animals are unlikely to be able to do: it is probable that they have a sense of the continuity of the self inasmuch as survival is its own testament to continuity, but it is a trivial sense which only serves to inform the current self. The trick of seeing time as episodic through the eyes of past and future selves requires a mechanism for identifying those past and future selves. Non-humans can see time as passing before the self, but they cannot see the self as passing through time.

To be human means being able to use second and higher orders of intentionality; it means being able to see others as mental agents, with a cognitive life as rich as the self's; it means being aware of my own mental life as a metacognitive event – being able to think about my thinking; it means being aware that there is a *me* to be thought about, to be planned for, and to have unrealistic expectations about; and it means having the ability to create a model world inside my head which is as significant to me as the real world outside my head.

All of these features rely on an ability to make models of myself inside my own head. This is a very unusual talent, and problematic in Darwinian terms: to make models of myself I have to step backward from myself: I have to try to view the “real” me from a fourth-person viewpoint. This means I have to be disinterested about myself, to try to see myself as others see me; and this is a skill that we are far from practiced at. Our self models are almost invariably wrong in significant ways: we overestimate ourselves and delude ourselves about our abilities. There is one person in the Universe that we need to be totally honest about, and we cannot do it.

---

<sup>19</sup> **Paul Bloom**, *How Children Learn the Meanings of Words*, ch3

<sup>20</sup> **Alison Gopnik, Andrew Meltzoff & Patricia Kuhl**, *How Babies Think*, pp155-162

Seen from this viewpoint, self awareness would seem to be a counter-productive developmental feature. And if a developmental feature is disadvantageous to the self then, in order for it to survive in a Darwinian world, it has to provide a larger, countervailing advantage. If the relative reproductive success it brings is great enough then almost any individually negative feature can be tolerated – including personal sterility, as in the *hymenoptera*. So what is the strong countervailing advantage that self modelling gave in order for it to become a standard feature of human individuals? That is the elephant in the room that has not been explored in this paper; but, like the hymenoptera, the advantage is likely to be intimately tied to our capacity for socialisation.

## Bibliography

- Roy F Baumeister, Jennifer D Campbell, Joachim I Krueger & Kathleen D Vohs.** 2005. Exploding the Self-esteem Myth. *Scientific American Mind* Vol 16 No 4, 2005, pp50-57.
- Paul Bloom.** 2002. *How Children Learn the Meanings of Words*. MIT Press: Cambridge, USA.
- Jerome Bruner.** 1986. *Actual Minds, Possible Worlds*. Harvard University Press, USA.
- Noam Chomsky.** 2000. *New Horizons in the Study of Language and Mind*. Cambridge University Press: Cambridge, UK.
- Richard Dawkins.** 1989. *The Selfish Gene*. Oxford University Press: Oxford, UK.
- Robin I M Dunbar.** 2004. *The Human Story: a new history of mankind's evolution*. Faber & Faber: London, UK.
- David Dunning, Chip Heath and Jerry M Suls.** 2005. Picture Imperfect. *Scientific American Mind* Vol 16 No 4, 2005, pp20-27.
- Martin Edwardes.** 2003. I like Both Myself and Me. In *Camling 2003: proceedings of the University of Cambridge first postgraduate conference in language research*. CILR: Cambridge, UK.
- Susan Greenfield.** 2000. *Brain Story: unlocking our inner world of emotions, memories, ideas and desires*. BBC Worldwide: London, UK.
- Marc D Hauser, Noam Chomsky & W Tecumseh Fitch.** 2002. The Faculty of Language: what is it, who has it, and how did it evolve? In *Science* vol 298 22 November 2002, pp1569-1579.
- Wilhelm von Humboldt.** 1836. *On Language: on the diversity of human language construction and its influence on the mental development of the human species*. Cambridge University Press: Cambridge, UK.
- James R Hurford.** 2003. The Neural Basis of Predicate-Argument Structure. In *Behavioural and Brain Sciences* 2003 26 (3), pp283-284.
- Julian Jaynes.** 1990. *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Penguin: London, UK.
- Kenan Malik.** 2000. *Man, Beast and Zombie: what science can and cannot tell us about human nature*. Phoenix: London, UK.
- Steven Pinker.** 1997. *How the Mind Works*. Penguin: London, UK.
- Ian Stewart & Jack Cohen.** 1997. *Figments of Reality: the evolution of the curious mind*. Cambridge University Press: Cambridge, UK.
- Michael Tomasello.** 1999. *The Cultural Origins of Human Cognition*. Harvard University Press: Cambridge, USA.
- Michael Tomasello & Josep Call.** 1997. *Primate Cognition*. Oxford University Press: Oxford, UK.