Connexions

An online journal of cognitive science

ISSN 1368-3233

EDITOR: TOM SIMPSON

Issue 5 March 2001

Contents

Articles

On the Origin of Symbols **T. E. Dickins** **2**

Connectionism is Nothing but Control Theory **Asim Roy** **19**

Reviews

George Dyson, *Darwin Among the Machines* **Tom Stafford** **24**

Sponsored by the Hang Seng Centre for Cognitive Studies

On the origin of symbols

Thomas E Dickins
Division of Psychology
Nottingham Trent University

Abstract: *Through a synthesis of Dunbar's (1993, 1996) and Gomez's (1998a, b) hypotheses on the evolution of language a further hypothesis about the origins of symbolic communication is made that relies upon simple learning. The aim of this speculation is to propose a specific origin story for symbolic communication and to marry behaviourist and cognitivist concerns about language. It is argued that this order of approach will enforce a more realistic parsimony on future models of language evolution.*

Introduction

This paper is about the possible role of learning in the origins of symbolic communication. As such it is a paper contributing to the literature on the evolution of language. Language evolution has recently become a distinct area within the broader push to integrate the behavioural sciences with evolutionary theory (Hurford et al, 1998; Knight et al, 2000). But what is the purpose of studying the evolution of language? Oliphant has recently commented on the role of such theorising:

> To theorise about the evolution of human language is to theorise about how human communication differs from the communication systems used by other species, and what biological basis underlies these differences. The features of human language that I would suggest we need to account for are as follows:
> **Syntax:** Human language is compositional, conveying structured meanings through the use of structured forms.
> **Learning:** Human language is passed on from one generation to the next via cultural transmission.
> **Symbolic reference:** The mapping between basic lexical elements and their meanings is arbitrary and conventional. (1998:1)

It is the latter two features that this paper focuses upon. Oliphant refers to specific <u>languages</u> being passed on generation to generation, through cultural transmission (see Tomasello, 1999 on the cultural origins of cognition). My interest is in the initial role of learning in the establishment of proto-symbolic communication, which I argue is the first significant stage in the origin of <u>language</u> per se. It is likely that once such communication was established it would be open to the refinements of cultural transmission, howsoever that is characterised, which in turn would lead to different <u>languages</u>.

Oliphant's statement of purpose also emphasises the role of evolutionary theorising in understanding the biological basis of contemporary linguistic behaviour. This can be read as a fairly broad claim. Evolutionary theories can help us to understand what other systems are involved in a capacity and how essential, or not, they might be in the formulation of a new one. This is in part due to the view of evolution by natural selection as a tinkerer adding on new bits when necessary and co-opting old systems for new problems where possible, as in Darwin's discussion of preadaptation (exaptation - Gould and Lewontin, 1979). Within the field of language origins much emphasis is made upon the probable social behaviours, communication systems and general cognitive and learning abilities of our pre-linguistic ancestors. What essential ingredient, or set of ingredients, had to be added to this mix in order to produce linguistic behaviour? How could such an ingredient have emerged? Or could some novel organisation of extant abilities have led to newly formed behavioural phenotypes? All of these putative explanations have biological bases in terms of neuro-physiology and also in relation to the underlying genetic component that produces such phenotypes.

Much of modern Evolutionary Psychology has been concerned with the origins of specific cognitive and behavioural abilities (cf. Barkow et al, 1992; Pinker, 1997). Evolutionary Psychology is concerned to discuss the underlying proximate psychological mechanisms that produce particular behaviours. This is the key difference between modern approaches and Sociobiology. Sociobiology (Wilson, 1975) was, and is interested in evolutionarily grounded theories of proximate behaviour, but not in the nature of attendant psychological mechanisms. Some theories of language evolution take a traditional Sociobiological perspective (e.g. Dunbar, 1993, 1996) concentrating on the kinds of social behaviours that led to the complex behaviour of language. Others fall within the Evolutionary Psychology paradigm (e.g. Bickerton, 1995) and discuss the emergence of specific psychological mechanisms or abilities that led to the production of proto- and full linguistic behaviour. Whatever form a theory of language origins takes it will discuss the selection of appropriate phenotypes from available variance or some order of mutation that accounts for the onset of a brand new and adaptive phenotypic feature.

It still might be objected at this point that the best way to understand a complex ability such as language is to study its ontogeny, its every-day operations and its various disorders due to damage and genetic failings. An evolutionary model would not tell you more about language than the total of these approaches and is more likely to be critically dependent upon these sources for a description of the end-state of evolution - a fully functioning adult human. This last point is no doubt true to an extent, but the order of findings listed still does not achieve a complete picture of language. One reason is as follows. Cognitive science is founded on the Turing machine conception of cognition and the empirical method of observing the regularity of input-output relations. As Hendriks-Jansen (1996; and see Dickins and Levy, in press) has recently argued, any number of computational devices could be hypothesised that would fulfil the input-output regularities observed in cognitive science. This is a consequence of the computational hypothesis, *pace* Turing, that dominates cognitive science – theoretically the brain can run any algorithm. The job of a natural science is surely to uncover the actual or natural kinds of computation that produce the behaviours after specific inputs. Evolutionary perspectives can hypothesise about component parts that might contribute to such input-output governance. In a sense, and this is Hendriks-Jansen's point, an evolutionary scenario can posit limitations on theories of the kinds of processing a cognitive agent can perform. None the less, the operations of such component parts are open to similar criticism. At best, it would appear that an evolutionary approach narrows the focus of such theory building.

It might also be argued that a full account of language must be one that includes an evolutionary element as language undoubtedly evolved (Pinker & Bloom, 1990). As such every aspect of the "whole view" of language and linguistic behaviour must be amenable to an evolutionary account, and if it is not then some doubt must be cast upon the merits of the view.

A common criticism of evolutionary theories is that they are merely tall stories about the origin of a given trait. Such "just so" stories are not open to empirical investigation or falsification but they appear to be plausible none the less. It is entirely possible to generate such poor models but it is not guaranteed that all evolutionary models are "just so" stories. It is perfectly possible to produce a model that will lead to predictions about extant behaviour and in so doing open itself to testing. It is worth noting that there is no reason to believe that other approaches in the behavioural sciences are immune to "just so" story telling. Given the nature of Turing inspired cognitive architectures a strong adherence to a particular computational story is hard to defend. Any proposed computation and a large variety of others could offer the same behavioural evidence in support. Behavioural data is not always sufficient information for distinguishing between different models of the production of that same behaviour. None the less, a careful refinement of cognitive theories leading to a set of very specific predictions will obviate this problem. Evolutionary theory can do the same.

This paper is concerned with the evolutionary emergence of a particular form of communication. The purpose of hypothesising about its evolution is to constrain future and more detailed theorising. As such this paper represents a philosophical exercise designed to prime intuitions and subsequent theory with a particular notion of communicative behaviour.

Communication is about the transfer of information (see below). An actor transfers information and in so doing alters the behaviour of a reactor (Krebs and Davies, 1993). This manipulation may be deliberate, the actor wanting to attain some goal through the reactor's behaviour. For example, a sexual display advertises desirability (and intent) to a reactor in the hope that the reactor will behave sexually with the actor. Information can also be used to deceive as we see in the literature on tactical deception. Byrne (1995) relates an infamous example of this. He describes how a young baboon waited until a nearby and unrelated adult had achieved the difficult task of exhuming a root to eat. At this point the infant let out a distress call alerting his mother, who outranked the root-seeking adult. She appeared to take in the scene and see the adult off as if he had attacked and deprived the infant of his root. Meanwhile the infant ate the root. This is a classic example of apparent intentional and deceptive communication. Byrne is not opposed to the notion that such information manipulation can be acquired through classical or instrumental learning.

Communication does not have to be deliberate. It is possible to communicate information about one's state of mind through gestures and facial expressions that might not come under conscious control. Equally it is possible that organisms leave tracks that constitute information for any predator pursuing them. Again, learning may well establish such relationships. Given that communication is only the transmission of information that has some effect upon the reactor – i.e. upon the recipient of that information – communication can be construed quite broadly. All behaviour is prompted by some informational input (be it externally or internally produced) and all behaviour can in turn be construed as information about the organism displaying it. Such behavioural information can be used by other organisms or by the behaving organism itself. There is therefore a sense in which an organism can communicate with itself. The extended hand supplies proprioceptive feedback that tells the organism about the nature of the external world (see Vihman and Depaolis, 2000 for a discussion of the role of such perception in language acquisition).

Despite the above argument communication is generally thought of in terms of deliberate information transfer from one organism to another. In other words, it is regarded as a social behaviour. Stretching this concept to incorporate the above example of "accidental" facial expressions might not alarm too many theorists as facial expressions are regarded as indicative signals of mood and have an evolutionary significance that Darwin (1872) himself noted. There is good reason for this information to be read by conspecifics and as such an evolved, subpersonal communication system does not threaten many views of communication. However, there are qualitative differences between various uses of information. Those interested in the evolutionary origins of our own linguistic communication system are keen to stress its deliberate and sophisticated nature.

Language, as used by a fully functioning adult *Homo sapien*, is typically regarded in Chomskyan terms as a complex syntactic system that constructs and computes strings of symbols. Each symbol, or word, has some independent meaning but the combination of such symbols in syntactically organised strings (sentences) enables the production and comprehension of novel meanings. Much of psycholinguistic theory has been concerned with constructing a cognitive theory that explains the kind and ontogeny of the computational architecture required for production and comprehension. Accounts of the origins of language have been heavily influenced by such theories (see Hurford et al, 1998, Knight et al, 2000) and they appear to regard language as a "quantum jump" away from other forms of communication. A change of such magnitude requires a very specific evolutionary explanation that will propose either mutations (e.g. Bickerton, 1990; Crowe, 2000; Place, 2000a, b) or the onset of other (cognitive) subsystems to mediate intentional communication (e.g. Bickerton, 2000a, b). However, if behaviour more generally is regarded as a systematic response to information, we might be tempted to try and draw a smooth trend from the simplest forms of behaviour through to the most complex, namely human language.

Another way of phrasing this last point is as follows. Chomskyan theory has undoubtedly met with considerable success in describing the formal structure of language. From this perspective the intricacy of the system has been revealed and computational explanations based around this description. However, evolutionary explanations run the danger of being dazzled by the complexity of the finished article and arguing for extraordinary circumstances and mechanisms. Pinker and Bloom (1990) made a similar point when they exhorted theorists to look for gradualist explanations of the evolution of language, for only in this way could such an apparently designed system emerge. This is quite possibly true but Pinker's own output in Evolutionary Psychology (1994, 1997) has extolled a specifically computational cognitive account and it is this order of explanation that is seen in much language origins work (Dickins and Dickins, in press) and much Evolutionary Psychology (Dickins and Levy, in press). Surely if such computational architecture constitutes this aspect of cognition there is an evolutionary account required of the onset of computational architectures either more generally or very specifically within the social communicative domain? As Dennett (1995) has argued, it is likely that the general evolutionary trend has been from hardwired behavioural solutions, to the addition of simple learning abilities and then the emergence of complex representational architecture (this is the argument behind Dennett's Tower of Generate and Test).

In order to think about the onset of putative computational cognition we perhaps need to see how far toward linguistic communication a creature, that had only the simplest behavioural abilities, could be forced. This would give a measure of when we had no other option than to hypothesise the introduction of novel mental abilities such as possible computational devices. This requires an emphasis upon the function of communication, the environment in which communication occurred as well as upon the more complex modern "end state". The lesson from Dennett (1995) is that evolutionary theorising should be seen as imposing parsimony upon models of extant behaviour. This parsimony is the parsimony of a tinkering "designer" who makes do with material to hand as best she can and, of course, evolution through natural selection is just such a "blind designer". In other words, we should not simply focus on selection and emergence of novel phenotypes that are dependent upon mutation. We should instead realise that selection operates upon behavioural solutions within a given problem space. Surely, it is more likely that genetic innovation will only be necessary when the problem space exhausts extant behavioural plasticity or learning?

This kind of approach to evolutionary theorising might best be termed teleonomic thinking. In principle the selection of a novel genetically based innovation can happen at any time. However, most mutations are phenotypically non- or mal-adaptive. To rely on mutation arguments heavily makes for a very unlikely set of theories. It is therefore reasonable to think in metaphorical terms of a rational designer making the best of what she has to hand. Only when she has run out of choices must she resort to adjusting the genes.

As symbols (words) are one of the basic constituents of full, natural language, without which syntax would have nothing to operate over, a good place to focus a parsimonious first pass at language origins is here. The first question to ask is "could symbolic behaviour be classically or operantly learnt under the right conditions?" If it could then we can avoid reliance upon mutation. However, such a learning theory theory would have to account for why some individuals suddenly acquired the trick and others did not. It might be the case that some organisms are better able to learn the trick and this will establish a selection pressure and an opening for a possible Baldwin Effect (Baldwin, 1896).

A learning theory explanation of the origins of symbols would require the following, not insubstantial list to be dealt with:

1. A definition of a symbol in terms of information learnt;
2. An account of the nature of learning involved in relating a symbol to an object, event or state of affairs;
3. An account of the origins of the media used in generating symbols such as the neuro-muscular control involved in speech or hand gestures;
4. An account of the selection pressures (i.e. the adaptive problems) that saw to the correlation of, for example, speech sounds and objects, events and states of affairs;
5. An account of how this correlation persisted over evolutionary time.

Any non-learning theory, or cognitive account of symbol origin would not necessarily deviate too far from the above explanatory requirements. One key difference would be the need to incorporate a characterisation of word learning in modern infants together with a computational theory of the underlying mechanism that allows this to happen (although a Baldwin Effect might explain the rapid and apparently canalised learning in modern infants). So, both a learning theory and a cognitive theory could be consistently similar in evolutionary terms but only differ at the point where the selection of modern mechanisms of word learning is involved. For the cognitivist the speed of acquisition and the lack of direct training or example undermine any Skinnerian (1957) story of linguistic ontogeny (Chomsky, 1959). It is likely that any evolutionary account of the origins of a computational capacity to acquire words in a non-instrumental fashion will require slight twists that a Skinnerian account would not follow.

Another key difference between a learning and a non-learning theory account would be that a learning theory account would have no need to posit a representational architecture whereas a cognitive one would. A cognitive account of symbol origins would require an account of the onset of lexical representation and an attendant semantic representation system which would be related to the lexical system such that word meanings could be easily produced and comprehended. A learning theory account would simply describe associated behaviour. This could be accounted for under a connectionist architecture that would then invoke idiosyncratic patterns of activation for any given word. But such connectionist representations do not have a causal role in the production of behaviour, they are merely a by-product of input-output relations. A connectionist model of this order could, in a limited fashion, mark the meeting point of behaviourist and cognitivist approaches to the behavioural sciences.

A Statement of Aims

This paper has two distinct aims. First, I want to propose a reasonably straightforward hypothesis about the origin of symbols, a hypothesis that I hope will be amenable to some order of simulation work at a later date. This hypothesis will not address all 5 of the points listed in the Introduction as part of a complete learning theory account of symbol origins. None the less, it represents an inroad into this task and as such should be treated more as a conceptual thought experiment rather than a directly empirically testable hypothesis, as indicated in the Introduction.

The second aim is perhaps more ambitious and results from the preceding discussion. I want to see how far we can get in our speculations about language origins using fairly low level "mechanisms", such as simple associative learning, before we have to posit major cognitive or architectural changes (cf. Balkenius et al, 2000; Place, 1996, 2000a, b). Such major changes would doubtless require attendant changes in brain organisation of some order. Current neurological evidence indeed suggests specialised brain regions involved in speech production and comprehension, but this is in the brains of fully linguistic creatures. What is more, the functional demands of a task such as language would undoubtedly lead to organisation of brain activity and this would not indicate an evolved set of brain regions necessarily. Understanding how much of the trick can be learnt will affect the number and nature of proposed neural adaptations.

This latter aim is not to be read as the wish of a radical behaviourist. Rather, it is an attempt to marry behaviourism and cognitivism, albeit at a low conceptual level. It is also an attempt to instil some parsimony into my own speculations about the origin of symbols and symbolic behaviour - something that I take to be an important step in the evolution of modern human cognition. These speculations take the form of a synthesis of a number of ideas in the language evolution literature, with some analysis.

I shall continue with a clear statement of my hypothesis, then a brief discussion about what symbols are, followed in the main part of the paper by the principal points of argument that led me to develop the idea. This should be the most efficient way of making my argument, but I recognise that some of the detail will be lost.

My hypothesis is that:

> The use of ostensive systems might have acted to direct the attention of hominid ancestors to the object of vocalisations. First this might have occurred solely within the environment of vocal grooming, as postulated by Dunbar (1993, 1996), but once a flexible vocal system was capable of being directed to specific grooming interactions object discriminations would easily follow. The main impact of social grooming would have been vocal control, the main impact of ostension would be that of guiding other behaviours and rendering them communicative. Once firm associations between vocalisation and object, event or state of affairs had been established through ostension the specific ostensive act would be unnecessary. This would liberate the vocalisation, which would, to all intents and purposes, be arbitrarily linked to its referent. Other symbol properties of displaced reference and symmetry would follow.

The last lines of this hypothesis make reasonably clear what I mean by "symbol", however, this will bear further clarification. Symbols are arbitrarily and symmetrically related to their referent, and in this way symbols are afforded the property of displaced reference. The relation is arbitrary because the symbol has no direct causal relationship with the referent. For example, there is nothing "rabbit-like" about the word <rabbit>. <Rabbit> is attached to rabbits by social convention (Deacon, 1997; Knight, 1998). Furthermore, if someone points to an array of objects and says <rabbit> we can pick out the rabbit from this array. Equally, we can say <rabbit> when someone presents us with a rabbit. In this way the symbol <rabbit> is symmetrically related to its referent.

Given that this hypothesis is being proposed within the broader discipline of language origins this view of symbols is designed to fit with stories about the origins of words. None the less, words are to be regarded as a species of symbol that have properties related to their grammatical role in sentences, among other things. These developments I take to be secondary to the primary development of communicative and vocal symbolic behaviour - if you like I am focusing upon proto-words.

There are two broad problems to be resolved in any theory of symbol origin. The first is the issue of how symbols relate to other forms of animal communication with a mind to understanding the key stages of this transition from non-symbolic communication to symbolic. The second problem is that of explaining the arbitrary and symmetrical attachment of a symbol to its referent. This second problem is made more clear when we consider Armstrong's, Stokoe's and Wilcox's (1995) definition of a symbolic gesture as something comprised of conceptual structure (ultimately reducible to neurological activity) that gives meaning and substantive content that allows them to be shared. Upon perceiving a symbol one can "switch on" the appropriate concept, and vice versa such that a concept can be communicated by you with an appropriate symbol. The production of a symbol is also ultimately reducible to neurological activity.

The field of language origins obviously assumes that these two problems are related through the selective history of language evolution. My suggestion is that there is little material difference between some non-symbolic and symbolic communication systems, in other words signalling systems have some of the properties of symbols but not all of them. The key properties that non-symbolic systems do not have relate to the production of arbitrary and symmetrical reference. These properties mark the first stage in the transition to linguistic communication as they afford a displaced reference that in turn affords greater representational flexibility in communication (and cognition). Given this assumption the selective story to be told is not too complex for at least this part of the emergence of language.

We have already seen that communication is traditionally regarded as the transfer of information from one organism to another, with the result that the receiver's behaviour, or internal readiness to behave, is in some way altered. Some theorists (e.g. Catania, 1991) maintain a more manipulative definition that might bring us up sharp against a discussion of intentionality. For now I want to avoid such a discussion. This is in part because it is not central to the aims of this paper but also because I suspect intentionality emerges hand in hand with language, partly as a result of shared attention (cf. Baron-Cohen, 1995, 1999; Tomasello, 1999) which will be discussed later in terms of ostension.

Marc Hauser (1996) has given a taxonomy of communicable information (see Table 1). Signals and symbols are to some extent similar. Signals, such as the vervet monkeys' repertoire of alarm calls, indicate specific objects, events or states of affairs. In the vervet case signals indicate specific predators and subsequently stimulate appropriate behavioural responses in other vervets (Cheney and Seyfarth, 1985, 1988). Symbols also indicate specific objects, events and states of affairs and trigger behavioural responses. It would appear that signalling systems and symbolic ones can discriminate categories in the world in much the same way and at much the same level. None the less, the relationship between signal and referent is not symmetrical and is based on affective response so nor is it entirely arbitrary. It is the properties of arbitrary and symmetrical relationship between symbol and referent that have to be explained by any symbol origins theory. Given this overlap of function we might hypothesise that the transition of interest in our ancestry is from signal to symbol.

Information Type:	**Feature:**	**Example:**
Cues	Always on	Yellow & black stripes of wasp
Signs/Indexicals	Indicate presence of something	Footprints
Signals	Can be on or off	Alarm calls
Symbols	Displaced reference	Words

Table 1: Summary of Information Categories: *Cues, signals and symbols are all used in communication systems. Signs or indexicals can be used by one organism to learn its way around the environment - this is related to communication as it is an example of behavioural change through the acquisition of information; but it lacks the dyadic (or more) interaction to be truly classed as communicative behaviour. Signals and symbols can be seen as a species of sign (Place, 1996).*

Initial Conditions

We now have a clear idea of what is meant by symbols and also the hypothesis. We can now proceed with the argument. First we need to think about what were the initial conditions in which symbols emerged:

Language is open to many definitions, as a casual glance at any introductory textbook will attest, none the less we can make some readily agreed general statements about the nature of language. First, most (but not all) languages are spoken which strongly suggests that the vocal medium is the one in which language emerged (Locke, 1998). It is noteworthy that many Great Apes have vocal communication too, although these vocalisations do not appear to come under the same order of neurological/motor control as those of modern humans. Second, modern humans are a social species and communication is a social activity (Dunbar, 1996). As with other Great Apes our social groupings are hierarchically organised (Byrne, 1995; Whiten & Byrne 1998) and probably have been for all of our history and that of our hominid ancestors. Given this, language is likely to have emerged as a part of, and reflects this, specific social economy (Bickerton, 2000a, b).

A vocal means of signalling within a highly social group is likely to have been in place before the onset of symbolic communication. However, there are other basic abilities that will have been in place in this putative ancestral grouping. It is almost a given that these ancestors will have been able to learn through classical and operant conditioning. What is more, the level of discrimination they will have been able to make when learning will be at least at the basic level of categorisation (see footnote 2).

As suggested in the Statement of Aims, we might do well to regard our problem as that of the transition from signal to symbol. Furthermore, I also suggested that one of the aims of this paper is to marry behaviourist and cognitivist concerns in order to impose some parsimony on theorising in this area. Given this it seems reasonable to ask what role simple learning might have had in the onset of signalling. I shall first outline how classical conditioning (or expectancy learning) might be involved in cues and then in signals. These examples are best regarded as thought experiments, or "intuition pumps" (to use Dennett's term), rather than precise hypotheses about the origins of these systems.

The example of a cue given above is the yellow and black stripes of a wasp. Wasps give a sting that is a painful irritant, but it is not fatal to many species. For now we shall imagine that knowledge of wasp stings is not innately coded and stored in other creatures. It is likely that yellow and black stripes are perceptually very distinctive to many creatures, hence the persistence of this coloration scheme throughout time, and that such a scheme could be associated with the properties of the wasp. In this case its unpleasant, irritant properties that are best avoided. This association could be made through excitatory classical conditioning. The black and yellow stripes would act as a Conditioned Stimulus (CS), the sting or irritant is the Unconditioned Stimulus (UCS) and the avoidance of the wasp that that causes is the Unconditioned Response (UCR).

1. [UCS (sting) – UCR (avoidance)]
2. CS (stripes) – [UCS – UCR]
3. CS – CR (avoidance)

How might this simple model transfer to signalling? Cheney and Seyfarth’s (1985, 1988) work on vervets has shown that they have specific signals for specific predators. How might a specific call result in a specific behaviour in the listener?

First, we can imagine a situation whereby an ancestral vervet monkey sees predator A and emits a "yelp". Another vervet will hear "yelp" and then some short time later see A. Seeing A will trigger normal flight responses, such as hiding under a bush. If this auditory clue is regularly paired with a temporally close sighting of a predator the vervet monkey that hears "yelp" might start to hide before seeing A.

In this case we can envisage the "yelp" as a CS, the sighting as an UCS and the hiding as a UCR:

1. [UCS (sight of A) – UCR (hide under bush)]
2. CS ("yelp") – [UCS – UCR]
3. CS – CR (hide under bush)

The existing UCS – UCR pairing might be learnt too or might to some extent be hardwired.

From Cheney and Seyfarth’s (1988) experimental work we know that vervets are very sensitive to false callers. It would appear that vervets pay close attention to the characteristics of the caller as well as the call. One explanation is that an unreliable caller may not present sufficient trials to learn a CS – CR relation. Alternatively, the unreliable caller may actually have an inhibitory learning effect. If the vervet always signals "yelp" and nothing happens (after the "yelp" to predator A association has been learnt from other sources) this CS will actually signal an omission of stimulus and response.

Both of these explanations rely upon the ability to accurately distinguish callers. If this distinction could not be drawn then one would expect stimulus generalisation to lead either to degradation in the signal-to-signalled relationships, such that the system ceases to operate, or to treating unreliable signals as accurate which would in turn lead to energy costs for no benefit on certain occasions. This could well be intolerable. However, distinguishing callers is not an unproblematic solution. If the vervets can distinguish all callers it either means that each vervet has to learn the associations between specific call-types and predators for each caller, or that stimulus generalisation allows the similarities in calls (they are strongly similar) to reduce the learning workload. If stimulus generalisation allows the recognition of call-types across individuals the problem of unreliable signallers again raises its head.

This circular problem need not be that vicious, however, if we assume that the information that distinguishes callers is only necessary when there is some anomaly in the calling system. If a CS fails to predict an UCS reliably then the monkeys will look to some new features to determine when the CS will predict and when it will not. In a sense the monkeys learn a new, finer grained CS.

It seems entirely possible, then, that expectancy learning will allow for the results found by Cheney and Seyfarth, and there is no reason to assume a sophisticated representational architecture to control this form of communication.

Much of the vervet signaling system may in fact be hardwired, the result of selection for ancestral learning. Certain sounds may well be emitted when certain predators are seen, and these sounds may stimulate relevant response in listeners. However, such calls and their responses are likely to be fine-tuned by expectancy learning (Oliphant, 1998). We can argue for a disposition to behave in such a way. Furthermore, it is possible that if the vervet repertoire is entirely hardwired it still had an ancestry like the learning account just outlined. Such a consistent learning environment might have led to the whole "system" becoming hardwired through Baldwinian means. The purpose of this expectancy account is to relate a possible scenario, not necessarily an actual one, which might help to explain the origins of human communication.

As I have already noted vervet communication is not symbolic, but it comes close. What is needed to make a new species of vocal signals that are arbitrarily and symmetrically linked to their referent?

The Transition from Signal to Symbol

Gomez (1998a,b) has recently proposed that the order of signalling systems seen in vervet monkeys might be "coaxed" into symbolic communication with the addition of an ostensive system that is under separate control. Gomez argues that an ostensive system would have allowed ancestral signallers to direct attention to new objects and attach vocal labels to them. He feels that the communication of vervets already exhibits a certain amount of "semanticity" and this simply has to be brought outside of the closed system of predator alarm calls - i.e. the referential base has to be extended. There are a couple of things wrong with this idea. First, it assumes a full referential status for vervet calls. This is not the case by my definition of symbols given that a referent has to be arbitrarily and symmetrically related to the "symbol" for it to be a symbolic relationship. In other words, full reference requires a symbol. At best the one-way call-to-behaviour relationship seen in vervet communication is a case of proto-referential behaviour. The second problem is that Gomez's speculation requires a significant amount of vocal control to enable the extension of vocal signalling into other domains.

From my hypothesis you will note that I am sympathetic to Gomez's idea and also that I have proposed marrying it with Dunbar's (1993, 1996) "gossip" theory of language evolution. Dunbar's basic idea is that as ancestral group sizes increased due to ecological pressures so time constraints grew with regard to physical grooming. Grooming was essential for group cohesion but it could not take up the whole day as food had to be found and other needs had to be met. Something had to replace physical grooming in this complex "society" and Dunbar suggests that proto-language of some form was favoured by selection for this role. I shall now explain this marriage of ideas in some detail. As you will expect this marriage is dependent upon the order of learning we have already discussed.

Dunbar's vocal grooming model is really a model about increased vocal control. Where physical grooming might once have been accompanied by murmuring, as we see in modern geladas, ancestral hominids had to increasingly rely upon this accompaniment instead of the actual physical grooming. Furthermore, in order for Dunbar's vocal grooming system to work, in terms of such issues as turn taking, initiating and ending bouts, etc. the use of ostension, possibly the eye contact variety that Gomez is keen to promote, may have become critical. Attention needs to be focused and directed - if you like, the intention of the vocal act has to be made explicit. Indeed, in modern Monkeys and Great Apes physical grooming often includes nudges and tugs etc. in order to get the "groomee" to position herself where the "groomer" requires.

Gaze direction would be a much more efficient in vocal grooming scenarios, but pointing might have had a role too. Gomez (1998a,b) has indicated a species difference in the use of eye contact as a way of directing attention. Monkeys do not have a flexible and free system of ostension through the use of eye contact – Great Apes do. This suggests that our hominid ancestors might have had such a free system too. It is undoubtedly the case that modern humans use eye contact. For instance, infants use the direction of gaze to ascertain the focus of attention in adult speakers (Locke, 1993). This often leads to the misallocation of names to objects, as adults do not always look at the objects they are speaking about. If this use of eye contact is a marked species difference it is likely that the ostensive system that Gomez is hypothesising is hardwired to some extent.

This ostensive clarification of vocal acts is nothing like the relationships within communication systems based on cues or signals — all of which have a direct causal link with the object, event or state of affairs they carry information about. Here the information conveyed by the vocalisation is tied to that vocalisation in an arbitrary manner, but on the back of the ostensive gesture. Once this link is made firmly enough, through simple learning, there is potentially no need for the accompanying ostensive gesture. The vocalisation can do the work on its own. If this is true, then we could also expect the utterance of a vocalisation learnt in this way to carry information about what the utterer intends the hearer to attend to, even if that thing is not present.

In order to clarify this last set of points we shall look at an imaginary ancestral scenario. Let us assume that that Gomez's speculation is correct and that vocal signalling of the sort seen in vervet monkeys, plus ostension leads to naming. Let us also allow for a few additions to this speculation:

- Vocal "semanticity" in the ancestral population is flexible - there is quite a lot of potential variance in the vocal output;
- The connection between the referent and its associated behaviour is not established on the back of an affective response, thus rendering it fairly flexible;
- Basic categorical perception constraints already exist - for example, the whole object bias, as the whole object has to represent a fundamental level of interaction between agent and world.

In this hypothetical situation the ancestral species has a behaviourally fixed repertoire of calls that are learnt. An average call will emerge from this learning for each referent. However, this average call has not become hardwired - learning is sufficient. With the onset of ostensive behaviour within this ancestral population it is possible for some of the potential vocal variance to be directed to refer to new things - especially if this happens before the average calls are fully learnt. Ostension allows individuals to redirect attention and establish new vocal references. A novel sound can be made, whilst attention is directed to a previously "un-named" whole object. Eventually that whole object is associated through simple learning with the novel sound and the ostensive guidance is no longer required.

What sort of selection pressures would lead to this? As it stands the pre-ostension system is shaped by the existing constraints of the environment that have meant a limited set of calls is sufficient signalling. New predators might make for new alarm calls, but it is improbable that such novelty would suddenly arise. In fact, the environment of our ancestors is likely to have been a stable savannah one for the last 7 million years. So what would possess a suitably vocally plastic ancestor to start finessing that variance and making new reference? Let us look at some more imagined scenarios:

Dunbar’s vocal grooming would not have occurred suddenly, but instead would have emerged over some time. As he has indicated it would be likely that certain vocalisations were already accompanying physical grooming and that there would be some variations within these vocalisations during different parts of the grooming interaction.

As time budget pressure increased due to increase in group size it might be the case that more emphasis was placed upon the accompanying vocalisations. Such vocalisations would act to mark various aspects of the grooming interaction. However, as more than two individuals would be involved when time budget constraints reached their critical limit certain key changes would have to occur within the vocalisations, whilst the physical contact would diminish to a minimum in most social exchanges.

The basic role of vocalisation can be represented as a simple expectancy learning trick like so:

Stage 1: Physical grooming behaviour 1 (P1) – Response 1 (R1)
So,
UCS (P1) – UCR (R1)
Stage 2: Vocalisation 1 (V1) – [P1 – R1]
So,
CS (V1) – [UCS (P1) – UCR (R1)]
Stage 3: CS (V1) – CR (R1)

As such a grooming relationship is held to occur between more than two individuals at a time it is critical that attention is directed in order to allow such expectancy learning to take place. If there were a barrage of vocalisations then the opportunity to direct grooming at individuals would be lost. In order to ascertain the object of the vocalisation eye direction could be noted – if an individual is obviously attending to another then the object of vocal grooming should be easy to establish. In this manner the vocal interaction could be carved into discrete and functional units that would service the relationships held between the group of individuals. The eye contact would allow specific vocalisations to be applied to specific individual hominids.

Manual ostensive gestures might aid this process in the early stages of change over from physical to vocal grooming. Actual grooming manipulations might become stylised. For example, physical contact such as a brief touch might indicate that the groomer is about to groom. Later this might just be a pointing gesture, perhaps accompanying obvious eye contact. This would be expected if vocalisation gradually took precedence over physical contact.

It might be objected that turn taking and individually directed vocal grooming would undermine the time budget gains made by forfeiting physical grooming. This is not necessarily the case – for instance, such vocal grooming could occur whilst other activities are being engaged in – such as hunting in small groups. Also vocalisations are much more likely to be rapid and so more relationships could still be serviced faster. Critically it is unlikely that a mass broadcast approach to vocal grooming would evolve, as this would not allow primate hierarchies to exist in the format they currently do. It may be the case that grooming facilitates coalitions but how much you groom an individual compared with another also indicates and reinforces rank. Rank is critically important in primate groupings and would have to be maintained somehow. It is highly probable that the relative amount of vocalisations directed at an individual is indicative of her rank.

From the vocal grooming scenario described above we can see that we have a model whereby the ostensive system, described by Gomez, has been made to narrow the focus of expectancy learning mechanisms in order to facilitate Dunbar's grooming scenario. Such a model is obviously predicated upon increased vocal control and this will demand a theory of neural control and the underlying physiology.

Once such a system is in place within grooming it does not appear too far-fetched to assume that vocal discrimination of other objects might occur – similarly directed by ostensive behaviour. At this point Gomez's intuition about hominid ancestors with extant "vocal semanticity" might become more useful to us. Once vocal control is established the plasticity that is denied to vervet systems would be in place and enable the addition of new sounds for new discriminations. Once relevant associations between sound and object had been made, through the guidance of attention, ostension would not be necessary and the sound could truly represent the object. Such vocalisations would be arbitrary – in that there would not necessarily be any causal link between them and the object they represented (this is likely to be true in the grooming scenario too).

The Emergence of Symmetry

We now have a model of arbitrary vocal reference arising through social and ecological pressures. However, we have not accounted for the symbol property of symmetry. This shall be discussed next beginning with a behaviourist possibility and then moving on to some speculations emerging from computer simulation work.

Place (1995/6) has recently discussed symbols and operant learning. By his account symbols can be regarded as discriminative stimuli that require two operant responses to be learnt - that of producing the symbol when presented with an instance of its referent, and, that of selecting the referent when presented with the symbol (Place; personal communication). There is potentially an operant situation for our ancestors if the contingency is socially reinforced through peer interactions. Such a two-fold learning procedure would ensure symbolic symmetry. In other words:

Schedule 1) A → B
Schedule 2) B → A
Leads to: A ↔ B

Place has given some clues as to how this might happen. Notable is an emphasis upon the social element in distinguishing symbol learning. Symmetry could be instated through social reinforcement when the relevant symbol is produced to discriminate the symbolised object, and vice versa when presented with the symbol only. In other words, these discriminations are a product of a social interaction.

Place argues that any new vocal discrimination that is learnt would have to be learnt operantly as this is the only way to learn symmetrical discriminations. In the vocal grooming scenario existing physical stimulus-response pairings were associated with vocalisations such that the vocalisation formed an expectancy. So Place's modification would have it that there was a transition from ostensive guidance of vocal expectancy learning (within grooming scenarios) to operant learning of vocal discriminations.

The main problem with Place's suggestion is that it is hard to envisage a suitable scenario in which our ancestors would start to operantly train one another. None the less, this is not sufficient reason to abandon simple learning strategies just yet.

Hurford (1989) and Oliphant (1996, 1997) have executed simulation work that has directly addressed the issue of symmetry and its emergence from associative learning. Hurford (1989) has dubbed the property of symmetry a "Saussurean sign, a bi-directional mapping between a phonological form and some representation of a concept" (p. 187) such that it can be used both for transmission and reception.

Hurford discusses how a language acquirer will make isolated observations of transmission about x and isolated observations about the reception of a signal about x. The language acquirer will also make non-isolated observations of transmission and reception about x. In either situation the trick the language acquirer has to pull off is the marriage of the two observations and the realisation of the bivalent nature of the symbol. Hurford sees a successful communication as "any encounter between individuals where one (the transmitter), while mentally attending to a particular concept, carries out some observable act (which may be a gesture, a vocalisation or whatever), and another individual (the receiver), as a result of observing this act comes to attend to the same concept" (p. 191). It is this trick that Hurford sets out to explain.

It is possible for this trick to be carried out without the use of a Saussurean sign. Hurford shows how two individuals could communicate with one another such that each individual has its own code for a specific concept that it transmits but each individual knows the others' transmission codes and is brought to the appropriate concept anyway. To borrow from Hurford, Individual A might transmit gesture 1 to refer to concept x. Individual B might receive gesture 1 and think of concept x. However, in order to transmit and bring concept x to A's attention, Individual B might successfully use gesture 2. Here we see that there is communicative performance but it is underlain by a clumsy system. What Hurford really wants to explain is how a Saussurean sign, a symmetric symbol, can come to be a shared mental representation that co-ordinates this bivalence. Hurford wants to explain this as he feels a Saussurean strategy will be advantageous in more complex communicative scenarios than those outlined in the example above. The use of a Saussurean sign will be the use of a high order mental representation, then, that will stipulate just such a co-ordination. He sets about this task using computer simulations of hypothetical pre-linguistic conditions.

Three strategies are tested in Hurford's simulations – Imitator, Calculator and Saussurean. These strategies are not supposed to represent actual mechanisms abroad in past or present times, but rather represent at least possible strategies that are available to be selected. In other words, simulations are a fairly sophisticated form of intuition pump.

Imitators sample both transmission and reception behaviours in their immediate environment. On sampling transmission they imitate that behaviour and transmit it themselves. Equally they imitate reception. Calculators use averages of their samples of transmission as their basis for reception behaviour and vice versa. So for reception:

> Ah, I see someone giving signal X in response to object Y, so I will henceforth interpret signal X as drawing my attention to object Y.

And transmission:

> Ah, I see someone responding to signal X by attending to object Y, so henceforth I will use signal X to draw attention to object Y. (p. 210)

The Calculator strategy is meant to be more manipulative, or calculating, than that of the Imitator.

> Imagine a creature capable of learning how to manipulate other creatures, using intentionally produced gestures, and also learning how to respond profitably to actions of theirs. The Calculator treats the problem of how to interact with others without recourse to any notion that he himself may be like them. He learns to interact with them, both transmitting and receiving, as if they were aliens, of another species. It is as if a person learnt by experience that a certain growl by a lion meant that the lion was about to attack him, and that giving a certain shout tended to drive the lion away. (p. 211)

Both Imitator and Calculator strategists make no attempt to co-ordinate transmission and reception behaviour. If the population from which samples are being taken is perfectly co-ordinated then both the Imitator and Calculator will be perfectly co-ordinated without the requirement of some higher order representation. As the Calculator is averaging his samples this will make it more susceptible to ambiguities and lack of co-ordination in the population but unlike the Imitator, the Calculator will produce the optimal communicative behaviour given these constraints.

Hurford notes an interesting aspect of the Calculator strategy:

> The Calculator strategy is one strategy, albeit a very simple one, embodying the idea that learning a signalling system is not a special kind of learning, but rather the application of a general intelligent learning ability. In this way, the Calculator strategy contrasts with the Saussurean strategy…, which models learning in a way consistent with the internalisation of a specially linguistic (or more generally, semiotic) construct, the bi-directional Sign (or symbol) (p. 211)

The Saussurean strategy is like the Imitator in that it acquires transmission data in the same way, but differs as it derives reception behaviour from the sample of transmission behaviour. Such a strategist can only hope to co-ordinate transmission and reception behaviour internally by talking to himself, as Hurford puts it.

> Ah, I see someone giving signal X for object Y, so X must be the sign for Y; I'll accordingly build myself a two-way mapping between X and Y, for the use both in transmission and reception. (p. 212)

So the Saussurean strategist is a far more cognitive beast than both the Imitator and Calculator. The latter strategists seem to have more in common with the approach supported thus far. They seem to show how symmetry might emerge in a communicating population, whose transmission and reception is co-ordinated to some extent, through simple learning of both elements. Could the onset of one of these strategies be imagined in a pre-symbolic population?

By my scenario we are already in possession of low level communication. This systematicity could be aided by Imitators or Calculators – remembering Hurford's point about the generality of these forms of learning and also that this interaction, even in manual grooming, can be regarded as rudimentary communication.

Hurford put the simulated populations for each strategy through a number of the following generation cycles:

1. Birth;
2. Acquisition of the basis of signalling behaviour according to the three strategies outlined above;
3. Selection as parents of the next generation by a random choice weighted according to individuals' communicative and interpretative potential in relation to the population as a whole;
4. Death.

It was assumed that proficient use of the adopted communication system afforded direct fitness benefits. Every combination of strategist competition was run. Three different starting conditions were trialed – random "i.e. no apparent signalling system at all"; emergent, where "the initial population's average behaviour reflected an emerging connection between some objects and some signals, but by no means perfect communication"; and, perfect, where the population began with "a single, unambiguous set of object-signal correspondences".

Saussurean strategists were clear winners, achieving a higher reproductive potential. When Calculators and Imitators competed without Sausurreans, Imitators won every time. There appears to be an adaptive advantage to symbolic symmetry in that it affords a communication system with optimal co-ordination leading to fitness gains.

Oliphant (1997) was surprised by the absolute failure of the Calculator strategy in Hurford's simulations. In his own work Oliphant had managed to show Calculators as more effective strategists than Hurford had. The difference between the two sets of simulations was that Hurford based his upon non-overlapping generations. This means that no individuals from one population ever existed in the subsequent generation – whole generations died together. This had no effect upon Imitator and Saussurean strategies but it meant that Calculators were basing their optimal output upon "measurements" taken from the preceding generation. It would be very unlikely that a brand new generation would mimic the last if there were no generational contact.

Oliphant's own work has analysed the Calculator strategy, which he refers to as Bayesian, in more detail.

> Despite the success of the Saussurean strategy, there are a number of reasons to look for a better learning procedure. First, (there is) the requirement that signals be plentiful (and this) is one that may not be met in a given ecological scenario. Second, the *a priori* incorporation of the Saussurean sign as a part of the learning mechanism is less satisfying than the possibility of having it emerge as a consequence of the learning dynamic. (p. 69)

He continues by suggesting that:

> a learner could (instead of being a Saussurean) try to construct a communication system based on its expected utility for communication with the population. In other words, a learner should try to maximise its ability to communicate with the average member of the population. To do this it is necessary to a) create a send function that maximises the probability of being understood by the population's average reception behaviour, and b) create a receive function that maximises the probability of correctly interpreting the population's average transmission behaviour. (p. 69)

Oliphant demonstrated that such a strategy leads to increased performance for the population as a whole. However, Oliphant notes that in a more realistic simulation of communication this strategy does not work in a truly Bayesian fashion as the "observed reception behaviour is biased by the population's send behaviour" (p.73). This means that reception of infrequent transmissions of certain signals are under-represented and the probabilities used to calculate the strategist's own reception-transmission behaviour are not truly representative of the populations' activity. Some order of normalisation is required to adjust for this and Oliphant looks at a number of methods for performing this statistical control.

Oliphant ultimately settles on a form of Hebbian learning as the simplest and ecologically most believable method of achieving normalisation. Oliphant (1997) uses a connectionist simulation to test Hebbian normalisation. In the network used

> the input patterns are signals and the output patterns are meanings… A signal is represented on the input layer by turning on a single one of the input units. A meaning is represented on the output layer by turning on a single one of the output units. Associations between signal and meaning are represented by the bi-directional weights that connect each input unit to every output unit. (p. 78)

As Oliphant notes this bi-directionality of weights between input and output units effectively implements a Saussurean sign at this stage – thus the two strategies have begun to merge. Hebbian learning is implemented on this network by increasing the weight between input and output units if both units are active. If they are inactive nothing happens to the weight, but if either the input or the output unit (but not both) are active then the weight is decreased.

This formulation of Hebbian learning can be translated into informal terms. If signal and meaning match, the weight between the two is strengthened – i.e. the association between the two is increased. If signal and meaning do not match then the association is decreased. This procedure affects the normalisation that Oliphant argues is critical to compensate for non-comprehensive sampling. Oliphant argues that this is because the co-ordination between reception and transmission is imposed by the nature of the Hebbian network – all accurate matches are accounted for. Hebbian learning is a general procedure that can account for much learning including expectancy learning.

Oliphant's (1997) refinement of Hurford's Saussurean strategy leaves us in a position where we can pin symmetry to observational learning underlain by Hebbian principles. Oliphant suggests that *Homo sapiens* are the only extant species that can learn observationally and as such the emergence of this form of learning was a considerable move toward the development of language.

Cheap and Honest Symbols?

The thrust of this paper has been the defence of a learning theory theory of symbol origins. This has taken into account various extant theories and attempted to unpack them in simple learning terms. However, in doing this we have disregarded a key issue in all models of language evolution, an issue that has been discussed in detail by Power (1998, 2000) in response to Dunbar's theory. This is the issue of costs of communication. I shall describe and deal with this problem in this penultimate section.

Primate work has supported the Machiavellian Intelligence Hypothesis (Byrne and Whiten, 1988; Whiten and Byrne, 1998). This hypothesis states that modern human intelligence has arisen from a manipulative social intelligence. This social intelligence is essential to existing within a highly structured group where rank is crucial to opportunity. If you are of the highest rank then you have to constantly "watch your back", if you are of a lower rank you have to learn to form coalitions and to cheat sometimes in order to better your situation. Cheating incurs costs if caught and this has to be weighed against any putative benefit. For example, if you get caught social censure may see to your exclusion from the group and this will decrease your fitness. Thus a balance must be struck between gaining the advantages of living in a social group and further increasing inclusive fitness. Given this constant balance, and the fact that conspecifics will cheat if they feel they can, which means others will incur the cost for the cheat's benefit, individuals are on the look out for cheats. This increases the difficulty for cheats.

Zahavi (1975, Zahavi and Zahavi, 1997) has argued that in order for an actor to convince a reactor, of the reliability or honesty of a communication some obvious cost has to be incurred. If communication is very cheap then the opportunity for cheating is clear. Misinformation can be spread at little to no cost to the communicator, and potentially to their great benefit, remembering that communication is designed to alter the behaviour of a reactor or receiver. This makes for an unstable system and communication will surely break down. The imposition of an obvious cost, however, is likely to demonstrate that the actor is engaging in honest communication. As Noble (2000) puts it "signallers sacrifice some of their fitness… in order to produce signals that will be believed by receivers".

Dunbar (see above) hypothesises that vocal grooming emerged to replace physical grooming due to time costs increasing beyond acceptability for our ancestors. Power (1998) has criticised this and claimed that this hypothesis removes, as a part of the argument, any notion of cost in vocal signalling – physical grooming was too costly in terms of time so a time saving procedure replaced it. Power (1998, 2000) suggests an alternative - a separate system that signalled a commitment to alliances with the recipients of vocal "communication" could have coevolved with vocal grooming in order for it not to fall foul of Zahavi's prediction. Power then looks at various costly rituals that support coalition and alliance formation in human groups.

Power is undoubtedly correct given Zahavi's "handicap principle". However, it is not at all clear whether or not vocal communication does incur costs. Possible costs exist. For example, it is well documented that vocalisation in other species, such as song birds, is extremely energy expensive. A great deal of food has to be consumed to support such a system. Talking is also energy expensive, as any lecturer will attest. But one might doubt how apparent this is to an onlooker or reactor, especially one with no empathic capability. None the less, the putative costs of vocal communication appear to demand investigation in order to render a learning theory account plausible. That the costs have not been determined is not a sufficient reason to abandon this approach just yet. When this future work is executed it is worth noting that humans do use language to deceive one another in modern populations. Confidence tricks are played using language to convince people of certain "facts" in order to relieve them of their money – benefit to the actor, cost to the reactor – which, in itself, fits well with definitions of communication. As a species we are peculiarly aware of this and look to other signals in an individual such as their "body language". We find it easy to dismiss someone as untrustworthy who never meets our eye during conversation and this, just possibly, is related to ancestral uses of gaze as hypothesised in this paper and Gomez's work. We also look to the actual content – is it consistent, does it sound plausible – and the manner of speech - is it said with an overly authoritative tone, is it too fluid, as if rehearsed? Given the complexity of our response to the vocalisations of our conspecifics it is just possible that our communication relies upon a measured distrust, and perhaps long term observations of people, rather than overt signals of cost.

Conclusion

This paper had the twin aims of outlining a specific hypothesis about the origins of symbols and at the same time constraining cognitive speculation through a consideration of simple learning and its role in the emergence of symbols. The former aim has been met through a marriage of Dunbar's and Gomez's hypotheses about language origin, and the latter by wedding these two theories with simple learning mechanisms. What has not been discussed in any detail is the underlying neural and physiological changes for flexible vocalisation or the exact reasons for transferring vocal grooming to vocal object discrimination - this awaits future work, along with a detailed analysis of potential costs. For now the hypothesis presented here should, as with all evolutionary speculation, be taken as constraining problem space a little until it is undermined. However, although the precise hypothesis might be falsified I hope that the general approach is worthy of some attention.

Acknowledgements

I should like to thank David Dickins, Jim Hurford and Mick Perkins for their useful comments about my hypothesis. I should also like to express my debt to Ullin T Place who discussed many ideas about behaviourism and language with me. Ullin died in January 2000.

References

Armstrong, D F, Stokoe, W C & Wilcox, S E (1995) Gesture and the Nature of Language Cambridge: Cambridge University Press.

Baldwin, J (1896) A New Factor in Evolution? The American Naturalist, 30 (June), 441 - 451

Balkenius, C, Gärdenfors, P & Hall, L (2000) The origin of symbols in the brain In: Desalles, J-L & Ghadakpour, L (Eds.) Proceedings of the 3rd International Evolution of Language Conference, Ecole Nationale Superieure des Telecommunications, 13 - 17

Barkow, J H, Cosmides, L & Tooby, J (Eds.) (1992) The Adapted Mind: Evolutionary Psychology and the Generation of Culture Oxford: Oxford University Press

Baron-Cohen, S (1995) Mindblindness: An essay on autism and theory of mind London: MIT Press

Bloom, P & Markson, L (1998) Capacities underlying word learning. Trends in Cognitive Science, 2 (2), 67 - 73

Baron-Cohen, S (1999) The evolution of a theory of mind. In: Corballis, M C & Lea, S (Eds.) The Descent of the Mind Oxford: Oxford University Press

Bickerton, D (1990) Language and Species Chicago: Chicago Press

Bickerton, D (1995) Language and Human Behaviour London: University College London Press

Bickerton, D (2000a) Foraging versus social intelligence in the evolution of protolanguage In: Desalles, J-L & Ghadakpour, L (Eds.) Proceedings of the 3rd International Evolution of Language Conference, Ecole Nationale Superieure des Telecommunications, 20

Bickerton, D (2000b) How Protolanguage Became Language. In: Knight, C, Studdert-Kennedy, M & Hurford, J R (Eds.) The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form Cambridge: Cambridge University Press

Byrne, R (1995) The Thinking Ape Oxford: Oxford University Press

Byrne, R & Whiten, A (1988) Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans Oxford: Oxford University Press

Catania, A C (1991). The Phylogeny & Ontogeny of Language Function. In: Krasnegor, N A, Rumbaugh, D, Schiefelbusch, R L & Studdert-Kennedy, M Biological & Behavioural Determinants of Language Development Sussex: Lawrence Erlbaum Associates

Cheney, D L & Seyfarth, R M (1985) Vervet monkey alarm calls: manipulation through shared information? Behaviour,94, 150 - 166

Cheney, D L & Seyfarth, R M (1988) Assessment of meaning & the detection of unreliable signals by vervet monkeys Animal Behaviour, 36, 477 – 486

Crowe, T J (2000) Did *Homo sapiens* speciate on the Y-chromosome? Psycoloquy, 11 (001) <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?11.001>

Darwin, C (1872/1998) The Expression of Emotions in Man and Animal.

Deacon, T (1997) The Symbolic Species: The co-evolution of language & the human brain London: Allen Lane: The Penguin Press

Dennett, D C (1995) Darwin's Dangerous Idea London: Allen Lane: The Penguin Press

Dickins, T E (2000a) Signal to symbol: The first stage in the evolution of language Ph.D. Thesis (University of Sheffield)

Dickins, T E (2000b) A non-modular suggestion about the origin of symbols. In: Desalles, J-L & Ghadakpour, L (Eds.) Proceedings of the 3rd International Evolution of Language Conference Paris: Ecole Nationale Superieure des Telecommunications, 82 - 86

Dickins, T E & Dickins, D W (in press) Symbols, Stimulus Equivalence and the Origins of Language Behaviour & Philosophy

Dickins, T E & Levy, J P (in press) Evolution, development and learning – a nested hierarchy? In: French, R & Sougne, J Connectionist Models of Evolution, Learning and Development London: Springer – Verlag

Dunbar, R I M (1993) Coevolution of neocortical size, group size & language in humans Behavioural & Brain Sciences, 16, 681 - 735

Dunbar, R I M (1996) Grooming, Gossip & the Evolution of Language London: Faber & Faber

Gomez, J-C (1998a) Ostensive behaviour in great apes: The role of eye contact. In: Russon, A E, Bard, K A & Taylor Parker, S (Eds.) Reaching into Thought: The Minds of the Great Apes Cambridge: Cambridge University Press

Gomez, J-C (1998b) Some thoughts about the evolution of LADS: With special reference to TOM & SAM In Carruthers, P & Boucher, J (Eds.) Language & Thought: Interdisciplinary themes Cambridge: Cambridge University Press

Gould, S J & Lewontin, R (1979) The Spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme Proceedings of the Royal Society B, 205, 581-598

Hauser, M D (1996) The Evolution of Communication London: MIT Press

Hendriks-Jansen, H (1996) Catching Ourselves in the Act London: MIT Press

Hurford, J R (1989) Biological evolution of the Saussurean sign as a component of the language acquisition device Lingua, 77, 187 – 222

Hurford, J R, Studdert-Kennedy, M & Knight, C (Eds.) (1998) Approaches to the Evolution of Language Cambridge: Cambridge University Press

Knight, C, Hurford, J R & Studdert-Kennedy, M (Eds.) (2000) The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form Cambridge: Cambridge University Press

Krebs, J R & Davies, N B (1993) An Introduction to Behavioural Ecology -Third Edition Oxford: Blackwell Science

Knight, C (1998) Ritual/speech coevolution: a solution to the problem of deception In: Hurford, J R, Studdert-Kennedy, M & Knight, C (Eds.) Approaches to the Evolution of Language Cambridge: Cambridge University Press

Leavens, D A & Hopkins, W D (1998) Intentional Communication by Chimpanzees: A Cross-Sectional Study of the Use of Referential Gestures Developmental Psychology, 34 (5), 813 – 822

Locke, J L (1993) The Child's Path to Spoken Language London: Harvard University Press

Locke, J L (1998) Social sound-making as a precursor to spoken language In: Hurford, J R, Studdert-Kennedy, M & Knight, C (Eds.) Approaches to the Evolution of Language Cambridge: Cambridge University Press

Mervis, C B & Crisafi, M A (1982) Order of acquisition of subordinate-, basic-, & superordinate-level categories Child Development, 53, 258 - 266

Noble, J (2000) Co-operation, Competition and the Evolution of Prelinguistic Communication In: Knight, C, Hurford, J R & Studdert-Kennedy, M (Eds.) The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form Cambridge: Cambridge University Press

Oliphant, M (1996) The dilemma of Saussurean communication BioSystems, 37, 31 – 38

Oliphant, M (1997) Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language Ph.D. Thesis <http://www.ling.ed.ac.uk/~oliphant>

Oliphant, M (1998) Rethinking the language bottleneck: Why don't animals learn to communicate? Paper presented to the 2nd International Conference on the Evolution of Language, unpublished MS form available at <http://www.ling.ed.ac.uk/~oliphant>

Pinker, S & Bloom, P (1990) Natural Selection and Natural Language Behavioural and Brain Sciences, 13, 707 – 784

Pinker, S (1994) The Language Instinct London: Penguin

Pinker, S (1997) How the Mind Works London: Allen Lane: The Penguin Press

Place, U T (1996) Symbolic Processes & Stimulus Equivalence Behaviour & Philosophy 23/4, 13 - 30

Place, U T (2000a) The Role of the Hand in the Evolution of Language Psycoloquy, <http://www.ai.univie.ac.at/archives/Psycoloquy/2000.V11/0012.html >

Place, U T (2000b) From icon to symbol: An important transition in the evolution of language Proceedings of the 3rd International Evolution of Language Conference, Ecole Nationale Superieure des Telecommunications, 183 - 184

Power, C (1998) Old wives' tales: The gossip hypothesis and the reliability of cheap signals In: Hurford, J R, Studdert-Kennedy, M & Knight, C (Eds.) Approaches to the Evolution of Language Cambridge: Cambridge University Press

Power, C (2000) Secret Language Use at Female Initiation: Bounding Gossiping Communities In: Knight, C, Hurford, J R & Studdert-Kennedy, M (Eds.) The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form Cambridge: Cambridge University Press

Tomasello, M (1999) The Cultural Origins of Human Cognition London: Harvard University Press

Vihman, M M & Depaolis, R A (2000) The Role of Mimesis in Infant Language Development: Evidence for Phylogeny? In: Knight, C, Hurford, J R & Studdert-Kennedy, M (Eds.) The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form Cambridge: Cambridge University Press

Whiten, A & Byrne, R (Eds.) (1988) Machiavellian Intelligence: Social Expertise & the Evolution of Intellect in Monkeys, Apes & Humans Oxford: Clarendon Press

Wilson, E O (1975) Sociobiology: The New Synthesis Harvard: Harvard University Press

Zahavi, A (1975) Mate selection – a selection for a handicap Journal of Theoretical Biology, 53, 205 – 214

Zahavi, A & Zahavi, A. (1997) The Handicap Principle: A Missing Piece in Darwin's Puzzle Oxford: Oxford University Press

Connectionism is Nothing but Control Theory

Asim Roy
Arizona State University

Abstract: *The paper shows that connectionist systems are based on standard control theory notions. It does so by first examining the notion of a controller in any system and then establishing that many of the simpler connectionist learning methods - like back-propagation, adaptive resonance theory (ART), reduced coulomb energy (RCE), and radial basis function (RBF) - use controllers in them. The paper shows the existence of controllers in these methods and shows that these controllers may be (1) within the learning system itself, (2) outside of the learning system or (3) a combination of the two. By logical extension, more complex connectionist systems, ones that use these simpler learning mechanisms within them, are also in turn using controllers and are therefore based on control theoretic concepts. The analysis of these connectionist systems is performed purely on a logical basis, by a logical analysis of their conceptual structure, and has nothing to do with their implementation, whether by use of neurocomputers or other kinds of computers. In general, this analysis implies that control theoretic notions are applicable to developing systems similar to the brain and refutes the claim in connectionism that their methods do not embody standard control theory concepts, that they have introduced a qualitatively new set of concepts and mechanisms*

Keywords: Connectionism, controllers, control theory.

1. Introduction

It has been argued by connectionists (Rumelhart and McClelland [1986], Rumelhart [1989], Fahlman and Hinton [1987], Feldman and Ballard [1982], Grossberg [1982, 1987, 1988], Kohonen [1989, 1993] and many others) that their methods do not embody standard control theory concepts, that they have introduced a qualitatively new set of concepts and mechanisms. This paper argues that connectionist systems are indeed based on control theoretic notions. It makes this argument by first examining the basic control theory concepts and then showing that some of the simplest connectionist learning methods - like back-propagation, adaptive resonance theory (ART), reduced coulomb energy (RCE), and radial basis function (RBF) - are actually based on those standard control theoretic notions. More complex connectionist systems, such as those that are either similar in structure to those simple systems or are based on those simple systems, are also, by extension, based on standard control theoretic notions. This paper, therefore, refutes the claim of connectionism that their methods do not embody standard control theory concepts. This implies that standard control theoretic notions should be applicable to developing systems similar to the brain and in understanding how the brain works and learns (Roy [2000], Roy et al. [1995, 1997a,b]).

The paper is organized as follows. Section 2 characterizes the notion of control and controllers in any system. In section 3, it is argued that some of the simplest connectionist systems are actually based on standard control theory concepts. This is followed by a conclusion in section 4.

2. Standard Control Theory - What are controllers? What are their characteristics?

In general, most complex systems can be decomposed into various subsystems according to their functionality. Controller-based systems are characterized by the presence of one or more controllers in them, controllers that control one or more subsystems within the system. In other words, in a controller-based system, there might be one or more subsystems in control of other subsystems that are subservient to them in some sense. Thus, in this case, the controlling subsystem can be called the "master" subsystem.

The main function of a controller is to supply certain operating parameters to the subservient subsystem. And that can be done in a variety of ways. Perhaps some examples will clarify this notion. For example, many man-made devices are generally operated by humans - e.g. a car, an airplane, or a TV. In these overall systems, the car, the airplane or the TV is the "subservient" subsystem and the human is the controller, the "master" subsystem. The overall system consists of both the man-made device (the car, the airplane, or the TV) and the human. The human in these systems supplies the operating parameters to the subservient subsystems. For example, the human uses the accelerator of a vehicle to set its speed, its steering wheel to determine its direction of movement, and the buttons of a remote control to control a TV. In subsequent discussions, the subservient subsystem, without its controller, will simply be called the "system" and the subservient system plus the controller will be called the "overall system."

2.1 The argument against the notion of a controller in any system

Before proceeding further, one has to deal with the arguments against the very notion of a controller, because the notion of a controller in any system is very much disputed by many in brain-related sciences. Dealing with this issue also brings out a very important property of controllers. The standard argument against controllers runs as follows: The car, the airplane or the TV that is operated by a human is actually a feedback system. In a feedback system, a subsystem receives inputs (feedback) from the other subsystems, and these inputs (feedback) are then used to determine its output(s), its course of action. Thus these subsystems are completely dependent on each other (co-dependent) for their outputs and, therefore, there is no subsystem controlling another subsystem in these overall systems. Thus, the argument goes, it is not proper to characterize the human as the controller in any of the above-mentioned systems - the car, the airplane, or the TV – because the human determines the course of action based on information (feedback) from these subsystems. Thus, it is further argued, if one is intent on claiming that the human is the controller in these overall systems, then the system itself (the car, the airplane, or the TV) can also be claimed to be the "controller," the one that controls the human. So, by these arguments, there are no controllers (central or distributed) in any system, natural or man-made.

On the surface, the above arguments seem to be correct, because in a feedback system, as in the examples above, the so-called controller's actions depend on the state of the subsystem it is trying to control. So the output of any subsystem, including the controller, is a function of the inputs (feedback) received by it from the other subsystems. So one is back to ground zero in trying to characterize controllers and distinguish them from the subsystems they control.

2.2 Back to square one – So who is a controller? What are its characteristics?

This "no controller in any system" argument has to be dealt with first before proceeding any further. A fundamental characteristic of a controller, unlike the system that it controls, is that it can determine its course of action (its output signals, that is) without considering the state of the system it is trying to control. In other words, the controller can also operate in a non-feedback mode. That means, it can operate with other types of inputs for the generation of its control signals and it is thus not necessarily dependent on feedback from the system it is trying to control. So a controller, by its very nature, is capable of operating in a manner independent of the system it is trying to control. Some examples will clarify this notion. For example, a person who is operating a TV with a remote control can simply close his or her eyes and turn the TV ON or OFF at random or in some predetermined fashion without considering what is being shown on the TV. Thus turning the TV ON or OFF is no longer depended on what is on the TV; the TV no longer has any influence on the controller - that is, the person with the remote control. But the operation of the system itself (the TV) still depends on the ON/OFF signal from the controller, that can't be or has not changed. In the case of a heating/air conditioning system, the thermostat can be set to turn the heating/cooling system ON or OFF every minute, or on the basis of any other predetermined time intervals, without any consideration of the actual temperature. Similarly, a driver can operate a car in a non-feedback mode by driving with the eyes closed and ears blocked, cutting-off all feedback through those channels, although the risk of an accident would increase substantially.

Thus the basic nature of a controller in any system is that it can operate in many different modes, with a different set of inputs, and therefore it's operation is not necessarily based on feedback from the system it is trying to control. The controller can still send different operating instructions to the subsystem it is controlling, but those operating instructions no longer depend on feedback of information from that subsystem. In contrast, the basic system itself is still dependent on operating instructions from the controller for its operation, that can't be changed. This is the main distinguishing character of a controller. And this is the character that entitles it to be called the "master" subsystem, because it can send signals (inputs) to the basic system and operate it in any arbitrary manner, even though other inputs to the basic system have not changed. In other words, the basic system is "depended" on the controller for its operations, but there is no such dependency the other way round.

It is important to clarify here that "controllers" can exist in other types of systems; they are not limited to being components of continuous, interactive feedback systems only. For example, many "master" subsystems have no continuous contact with their "subservient" subsystems. For instance, consider a doctor who provides treatment and medication to a person who is ill. After a certain treatment or medication, the doctor may not have any contact with the actual biological processes of the patient for a while, unless there is some kind of continuous monitoring system used to monitor the patient. Suppose, the doctor monitors the condition of the patient daily and tries other medication or treatment if something doesn't work. In this situation, the doctor is not in continuous contact with the subservient system (the patient), but only has intermittent contact. The doctor, however, is still the controller in this situation, since he or she can be arbitrary in the treatment of the patient. So this is a type of system where the controller can be "out-of-contact" with the subservient system for some period of time. In other systems, the controller can be the source of a single, one-time signal or decision, instead of multiple signals or decisions being provided in continuous feedback systems. Examples of such situations include 1) starting a random fire in a forest, and 2) spreading a bad rumor that starts a riot in some place. One might claim that the controller in this case loses control of the subservient system. But despite the subsequent "out-of-control" character of the subservient system, there is no question about the controlling nature of the "master" subsystem that starts the process. So, the "master" subsystem in all of these different systems are still "controllers" because of the arbitrary way in which they can determine the signals being sent to the "subservient" subsystems, whether it is one-time or multiple times.

3. Is connectionism based on standard control theory notions?

To answer the question "Is connectionism based on standard control theory notions?" one has to establish that connectionist systems either have controllers inside them or use outside controllers or both. To establish this, all one needs to do is examine some of the simpler connectionist learning systems and show that they use controllers. Once that is established, it then follows by logical extension that more complex connectionist systems, those that in turn use these simpler learning mechanisms or are similar in structure, also use controllers, and thus are based on control theory notions. The logical analysis of the simpler connectionist learning systems is based on the following determination: (1) what decisions are being made by the individual subsystems, (2) whether those decisions are being conveyed from one subsystem to another, (3) whether those decisions can potentially be arbitrary, and lastly, (4) what decisions are being conveyed to these systems from the outside by external sources. In general, if a potentially arbitrary decision made in one subsystem A is conveyed to and used by another subsystem B, then it means that subsystem A controls subsystem B. The key test here for a subsystem to be characterized as a controller is its "potential to be arbitrary." In the normal mode of operation, a controller might very well operate on the basis of certain rules, but that does not negate the fact that it is still a controller with the "potential to be arbitrary." For example, the motor control portion of the brain operates on the basis of some rules acquired over time, but that does not mean that it can't control the various limbs of the body in an arbitrary manner without following those rules. Similarly, the central bank in a country might operate on the basis of some economic models and rules, but that does not mean it "potentially" can't set the interest rates in an arbitrary manner ignoring those models and rules, although that might never happen.

It would be appropriate at this time to examine a few simple connectionist systems and establish that they use standard control theory notions. Here some standard connectionist learning methods are examined and it is shown that they employ controllers in their operation and thus they, in turn, use standard control theory concepts. As stated earlier, larger and more complex connectionist systems, those that employ these simple learning mechanisms or are similar in structure to them, would also be using control theory notions, because the subsystems embedded in them do or because of their similarity in structure to these simpler systems. It might be appropriate to start with the best-known connectionist learning method: the back propagation algorithm (Rumelhart and McClelland [1986]). For algorithms like back-propagation, including any variations of it, an external agent (perhaps a human, perhaps another module in an overall connectionist system) supplies from the outside the design of a network and the values for the various learning parameters that are necessary for it to learn. Since the outside agent can determine the network design and the various learning parameters potentially in an arbitrary fashion (this arbitrary nature of the agent can be verified with certainty when a human is the external agent and is providing the information to back-propagation), the outside agent is therefore a controller for the back-propagation method. Thus the back-propagation learning method employs an external controller, much like a driver in a car or a pilot in an airplane. The back propagation algorithm, therefore, is based on standard control theory notions.

For other connectionist learning methods like adaptive resonance theory (ART) (Grossberg [1982, 1987, 1988]), the reduced coulomb energy (RCE) (Reilly, Cooper and Elbaum [1982]), the radial basis function (RBF) networks (Moody & Darken [1989]) and the like, the network design module is inside the algorithm. So where is the controller in them? The usual design task in ART, RBF, RCE and the like is to add a new prototype or exemplar (a center and a radius or width) to the network. The design task, therefore, is to expand the size of the network as and when necessary. The training task is to make adjustments to all those prototypes or exemplars – that is, to adjust their centers and widths or radii. So, logically, the design and training functions are housed in separate modules in these learning schemes. So, as before, the design module supplies to the training module the design of a network. So the logical structure of these algorithms is similar to back-propagation - the network design and training modules are decoupled and the design module supplies design decisions to the training module. Therefore, since the network design decisions and decisions about the various learning parameters come from one or more outside sources to the training module (the network design from the design module; the various learning parameter values provided by sources external to the algorithm, perhaps a human or another module within an overall system), these outside sources act as controllers to the training module in algorithms like ART, RBF, RCE and the like. Thus these learning methods (ART, RCE, RBF, and the like) employ one or more controllers (internal and external) in order to work. These learning methods, therefore, are based on standard control theory notions. Again, if these methods are embedded and used in more complex connectionist systems, then they too, by extension, are based on standard control theory notions.

For the algorithms discussed here, and for other connectionist systems with similar structures or connectionist systems that embed these algorithms within them, the interaction between the training module and the controlling source(s) (human outside agents, internal modules) is not just a one-time affair. In fact, the training modules of these algorithms come back to the controlling sources over and over again with their results and the controlling sources then, if not satisfied with the results, provide new network designs and new sets of parameter values and orders the training modules to get new solutions on that basis. This is much like a doctor trying a particular medication or treatment on a patient, waiting to see the outcome after such a treatment, and if the outcome is not satisfactory, trying a different medication or treatment and so on. So the interaction between the training modules and the controlling sources in these algorithms, like the doctor case, is not just a one-time affair; it happens continually. However, the controlling sources are out-of-contact with the training modules for the time during which they are getting a solution, much like the doctor and patient case.

Note that the use of control theoretic concepts is part of the conceptual structure of these learning systems, independent of how they are implemented and, in particular, whether or not a neurocomputer, with parallel computation capabilities, is used for implementation. Also note that the foregoing analysis of the learning methods was based strictly on their logical structure and nothing else.

4. Conclusion

This paper proposed to show that standard control theory concepts are used in connectionists systems. It has done this in two parts. First, it examined the notion of control and controllers in any system and characterized controllers as subsystems in an overall system that can operate in an arbitrary manner. Using this characterization, it then performed a logical analysis of the structure of some core connectionist learning systems - like back-propagation, adaptive resonance theory (ART), reduced coulomb energy (RCE), and radial basis function (RBF) - and showed that they have and do use controllers in them. Sometimes these controllers are housed within the learning system, sometimes they are external to the learning system (like back-propagation and the like), and sometimes it is a combination of the two (like ART, RCE, RBF and the like). But the paper clearly establishes that controllers do exist in these systems and thus these systems are based on control theory concepts. By logical extension, the more complex connectionist systems, the ones that use these simpler learning mechanisms inside them or rely on them in some other manner or are similar in structure, also use controllers and are, in turn, based on standard control theory notions. This refutes the claim in connectionism that their methods do not embody standard control theory concepts, that they have introduced a qualitatively new set of concepts and mechanisms. This implies that control theory concepts can indeed be used to construct brain-like systems and in understanding how the brain works and learns (Roy [2000], Roy et al. [1995, 1997a,b]).

References

Fahlman, S. E. and Hinton, G. E. (1987). Connectionists Architectures for Artificial Intelligence. Computer, 20, 100-109.

Feldman, J. A. and Ballard, D. A. (1982). Connectionists Models and Their Properties. Cognitive Science, 6, 205-254.

Grossberg, S. (1982). Studies of Mind and Brain: Neural Principles of Learning Perception, Development, Cognition, and Motor Control. Boston: Reidell Press.

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. Cognitive Science, 11, 23-63.

Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures. Neural Networks, 1, 17-61.

Kohonen, T. (1989). Self-organization and associative memory. 3rd ed. Berlin, Heidelberg: Spriger-Verlag.

Kohonen, T. (1993). Physiological interpretation of the self-organizing map algorithm. Neural Networks, 6, 895-905.

Moody, J. & Darken, C. (1989). Fast Learning in Networks of Locally-Tuned Processing Units, Neural Computation, 1(2), 281-294.

Reilly, D.L., Cooper, L.N. and Elbaum, C. (1982). A Neural Model for Category Learning. Biological Cybernetics, 45, 35-41.

Roy, A., Govil, S. & Miranda, R. (1995). An Algorithm to Generate Radial Basis Function (RBF)-like Nets for Classification Problems. Neural Networks, Vol.8, No.2, pp.179-202.

Roy, A., Govil, S. & Miranda, R. (1997a). A Neural Network Learning Theory and a Polynomial Time RBF Algorithm. IEEE Transactions on Neural Networks, Vol. 8, No. 6, pp. 1301-1313.

Roy, A. & Mukhopadhyay, S. (1997b). Iterative Generation of Higher-Order Nets in Polynomial Time Using Linear Programming. IEEE Transactions on Neural Networks, Vol. 8, No. 2, pp. 402-412.

Roy, A. (2000). On Connectionism, Rule Extraction and Brain-like Learning. IEEE Transactions on Fuzzy Systems, Vol. 8, No. 2, pp. 222-227.

Rumelhart, D.E., and McClelland, J.L.(eds.) (1986). Parallel Distributed Processing: Explorations in Microstructure of Cognition, Vol. 1: Foundations. MIT Press, Cambridge, MA., 318-362.

Rumelhart, D.E. (1989). The Architecture of Mind: A Connectionist Approach. Chapter 8 in Haugeland, J. (ed), Mind Design II, 1997, MIT Press, 205-232.

Review

George Dyson, *Darwin among the machine*

(1997) Allen Lane The Penguin Press, ISBN 0-713-99205-0, pp. 286, Hbk £20

Tom Stafford
Department of Psychology
University of Sheffield

Chapters: 1. Leviathan; 2. Darwin among the machines; 3. The General Wind; 4. On Computable Numbers; 5. The Proving Ground; 6. Rats in a Cathedral; 7. Symbiogenesis; 8. On Distributed Communications; 9. Theory of Games and Economic Behaviour; 10. There's plenty of room at the top; 11. Last and First Men; 12. Fiddling while Rome burns.

This review begins with the background to the book, a summary of the general themes and my impressions on reading it. The second half of the review covers in more detail the contents of the 12 chapters.

The title comes from Samuel Butler. In the second half of the 19th century Butler predicted that the proliferation of machines would lead to their evolution in symbiosis with mankind. This is the essential speculation of the book, but a speculation which takes in the historical, philosophical and technical antecedents of the neo-Darwinian synthesis, artificial intelligence, and the theory of informational systems in general.

Dyson is a kayak builder and ethnohistorian, self-educated in the libraries of the university campuses where he grew up. His family background is uniquely appropriate for writing this book on machines as minds, and minds as machination. His father, Freeman Dyson, is a mathematical physicist (who also made a foray into theoretical evolutionary biology with his book 'The Origins of Life'), while his mother, Verena Huber-Dyson, is a logician who wrote a book on the implications of Godel's Theorem. In his introduction Dyson says that observing his parents' work, and his sister's as a computer industry analyst, provided him with the foundations to write this book.

At the beginning of the book Dyson promises that this is a historical work on the nature of machines, and on predictions about them that turned out to be right. But he also asks some provocative questions about the prospects of man (and life in general) in the information age: 'Do we remain one species or diverge into many? Do we remain of many minds, or merge into one?' Across the 12 chapters of the book it becomes clear what Dyson's own answers to these questions are. His main themes are the dynamics of evolution (and signs of intelligence in this process) and the universality of intelligence in **all** complex adaptive systems. He concludes that a super-ordinate machine-human intelligence will inevitably evolve, or may in fact have already evolved in a form that we fail to recognise due to our anthropomorphic conception of intelligence.

The historical sketches are richly detailed, revealing Dyson's obvious warmth for the characters. Extensive quotations give a real feel for the style of the thinkers he covers, although one should be cautious about interpreting as prophesy the ramblings of centuries old thinkers. Dyson combines a flair for the biographical sketch with a passion for original ideas. Particularly memorable is the picture he paints of Lewis Fry Richardson, a meteorologist and mathematician who spent his evenings in the trenches during World War I calculating by a hand a grid of linked differential equations to predict weather systems - what amounted to a cellular automata, but done by hand rather than using the computer technology that has made effective exploration of these systems possible. More familiar figures fill the pages (e.g. von Neumann, Turing) as well as some familiar historical personages presented in unfamiliar context (e.g. the discussion of Hobbes in chapter 1 and his anticipation of artificial intelligence).

Ideas across the fields of evolution, mathematics, artificial intelligence and telecommunications are developed out of these sketches of the figures who first explored them. Readers who feel overly familiar with the history surrounding any particular theme – in my case it was the origin of the first computing machines, from Babbage through to Turing and von Neumann – should feel free to skip to the end of the chapters where Dyson pulls the threads together and gives full reign to his speculations. Reading *Darwin Among the Machines,* I could not help thinking that Dyson's self-education has left him particularly free of the normal blinkers that are acquired in a traditional education. There is a rich crossfertilisation of ideas from different fields and applied to all forms of informational networks: species, ecologies, brains and telecommunications networks. I find it difficult to express the sheer fecundity of the ideas Dyson has woven together in this book. In the chapter outlines below I have focussed on the main ideas and the ones which hold a particular interest for me. I wouldn't want any reader to think that I was presenting anything other than a biased and partial summary of the book. In particular I have focussed on the main characters of each chapter, rather rudely ignoring the many people Dyson typically discusses as he sketches the history leading up to any particular innovation.

1. Leviathan

For seeing life is but a motion of Limbs, the beginning whereof is in the principall part within; why may we not say that all Automata (Engines that move themselves by springs and wheels as doth a watch) have an artificiall life Thomas Hobbes

Thomas Hobbes, according to Dyson, "believed life and mind to be the natural consequences of matter when suitably arranged". From Hobbes' functionalism Dyson goes on to discuss the prospects for a second occasion when life might evolve from unthinking, inanimate matter. Ideas of collective intelligence, universal communication & symbiosis with computers are foreshadowed. Dyson says "Nature has begun to claim our creations as her own". In other words, the undirected, or at least unintentional, evolution of computers is taking off.

2. Darwin among the machines

Why not may there arise some new mind which shall be as different from all present known phases as the mind of animals is from that of vegetables? Samuel Butler

The feud between Charles Darwin and Butler is covered. Butler attacked Darwin for, amongst other things, stealing his grandfather Erasmus' ideas and failing to give him credit for them. Erasmus Darwin anticipated the essentials of Charles' theory of evolution by natural selection but, partly due to an obscure presentation style, he is usually associated with the errors of Lamarckianism and group selection. Dyson's claim is that neither Butler nor Erasmus Darwin suffered from these fallacies, but instead that they were grappling with, in early form, issues about self-organisation and emergent properties of systems. So that, when Butler wrote that species "very gradually, but nonetheless effectively, design themselves", he was hinting at an intuition that the mechanisms of evolution possess an intelligence beyond that of pure 'blind' chance.

"Butler's whole nature revolted against the idea that the universe was without intelligence" wrote H.F.Jones. His belief that there was a intelligence to the structure of universe anticipates views of those who believe that matter has an inherent propensity to self-organise, aside from (but complementary to) the forces of natural selection. Butler's understanding of evolution was far from naïve: he understood that evolution required some sort of mechanism like DNA long before it was discovered and his discussions on the evolution of ideas anticipate Dawkins' *meme* concept.

Dyson's father asked if there was a logical connection between metabolism and replication; proposing that the two separately developed and that the current genotype-phenotype replication mechanism is a symbiont with the Lamarckian development of metabolising (i.e. self-sustaining) cell types. Dyson (junior) then goes on to point out the relevance of Lamarckian evolution to the current – in other words, primitive – stages of machine evolution.

3. The General Wind

…that idea of some of the ancient writers…according to which souls are born when the machine is organised to receive it, as organ-pipes are adjusted to receive the general wind Gottfried Wilhelm von Leibniz

Leibniz's progress on formal logic and mechanical calculation inspired Babbage's analytical engine which, if realised, would have been the first machine to mechanise arithmetic. This use of limited symbols and rules (of some kind) to perform unlimited calculations is the obvious requirement for an intelligent system. In this chapter Dyson also covers an idiosyncratic history of the systemisation of the mental and formal systems, encompassing Boolean logic and Godel among others.

4. On Computable Numbers

Turing's work was obviously crucial to the development of AI. The universal Turing machine, with its two fundamental assumptions of discreteness of time and discreteness of states of mind, demonstrated that all computations are equivalent on some level (a good thing). More problematically, he also demonstrated the existence of non-computable functions.

Dyson comments "It is surprising that noncomputable functions, which outnumber computable ones, are so hard to find. It is not just that noncomputable functions are difficult to recognise or awkward to define. We either inhabit a largely computable world or have gravitated toward a computable frame of mind." Quoting Danny Hillis "In fact, it is difficult to find a well-defined example of a noncomputable function that anybody wants to compute. This suggests that there is some deep connection between computability and the physical world and/or the human mind."

Turing's dismissal of the use of Godel's theorem to argue against artificial intelligence was succinct "in other words then, if a machine is expected to be infallible, it cannot also be intelligent". Godel's theorem, Dyson continues, "demonstrated not a theoretical obstacle, but simply the need to develop fallible machines able to learn from their own mistakes". Turing saw that the key to intelligent machines was allowing them to alter their own programming – to learn or develop as we do from childhood. Like von Neumann he is usually associated with traditional, serial processing. However Turing recognised that intelligence was a collective property, something developing from the interaction of lesser parts, and that this went hand in hand with parallel processing.

5. The Proving Ground

The cold war was a proving ground for the development of high-speed electronic calculation (to aid nuclear bomb design) and, incidentally, game theory (to guide strategy). The war catalysed the development of the first general purpose calculator – the IBM 701, originally called 'the defence calculator'. As a side note it has been claimed recently that IBM's punch card technology, which was so important for the atomic project, was also instrumental in the Nazi orchestration of the holocaust (see "IBM and the Holocaust" by Edwin Black, 2001, Little, Brown & Co.).

6. Rats in a Cathedral

Was man indeed, as he sometimes desired to be, the growing point of the cosmical spirit, in its temporal aspect at least? Or was he one of many million growing points? Or was mankind of no more importance in the universal view than rats in a cathedral? Olaf Stapledon

The history of the Institute of Advances Study is covered. The IAS was home to many remarkable thinkers, including Einstein, Godel, von Neumann (with other notables passing through, such as T.S.Eliot, Piaget) in the 50s and 60s. There Julian Bigelow supervised the building of von Neumann's high speed general purpose computer – 'the IAS machine'. This used the famous von Neumann architecture, the success of which has obscured von Neumann's commitment to parallel computing as the necessary way to understand the brain and his belief that the interaction of smaller parts could generate emergent properties – namely intelligent behaviour – in the whole. These minor parts are the rats in the cathedral of the chapter's heading.

7. Symbiogenesis

The evolution of languages is a central mechanism by which life and intelligence unfold…Languages survive by hosting the reproduction of structures (letters, word, enzymes, ideas, books or cultures) that in turn constitute a system sustaining the language from which they sprang George Dyson

The IAS computer was used by Nils Aall Barricelli to develop a working model of evolution and investigate the role of symbiogenesis in the origin of life. Defined as evolution by the combination of simple forms, symbiogenesis seems crucial to the increase in complexity seen in evolution. Creationists commonly criticise evolution as a degenerative process, having no tendency to increased complexity. Superficially a simple mutation and selection model seems to lack the sophistication to overcome the powerful 2nd law of thermodynamics: the universal trend towards entrophy (energy dispersal). Barricelli's early experiments in simulated evolution allowed the observation of many basic 'biophenomena': sexual reproduction, self-repair, parasites and evolutionary stagnation. From these Barricelli concluded that mutation and selection alone were not sufficient to account for the speed of evolution. Sex and symbiogenesis are two crucial 'meta-evolutionary mechanisms' which aid the efficient parallel search of genetic space. Dyson provocatively notes that efficient search is a mark of intelligence, but does not develop the theme of the intelligence of evolution until later chapters.

Barricelli also focused on the importance of a complex environment for the evolution of complex forms. Stagnation and local minima can be avoided by building diversity into the environment by, for example, providing different environments in terms of different artificial rules for reproduction and mutation. This avoids a monoculture of 'organised homogeneity'. Thomas Ray's Tierra simulations also illustrate the importance of environmental diversity. In particularly the requirement of semi-permeable boundaries between diverse and changing environments. Total permeability creates unstable ecosystems that 'crash' upon the evolution of some new parasite, or tend towards dull monocultures - local minima from which no new evolutions can be successful. This relates to work by one of the founding fathers of cybernetics on 'connectance' within complex systems. Ashby found that beyond a certain level of connectance complex systems flipped from being stable to unstable. To my mind this suggests a deep connection between ecologies, various brain anomalies (e.g. epilepsy and schizophrenia) and economies. For example, the markets with the most connections, the freest information exchange and the smallest lag between cause and effect are the financial markets. Given their notorious instability it is worth giving a moment to ponder the wisdom of relentless and limitless trade liberalisation. Economic connection without any barriers may lead to the same type of monoculture (of monopolies?) as ecological connection without barriers leads to in the Tierra simulations.

In this chapter Dyson also starts to warm to another key theme of his – the importance of languages to relate between different levels of inter-linked systems. Thus the phenotype/genotype distinction is important because it allows translation between the gene and protein level. In a sense the genotype learns to 'operate' (exploit) the phenotype language (proteins). Gene code is selected due to its effects on the phenotypic level, with many genes able to code for the same outcomes. Coding into proteins allows a more effective search of evolutionary space, for the same reason that you are more likely to accidentally come up with a grammatical sentence if you choose words at random than if you choose letters at random. Computer code is another kind of self-reproducing information – which through an electronic expression might learn to control aspects of the world as the genotype has learnt to control organisms. The interaction of a-life (digital code) and b-life (biological code) increases the opportunity for human-computer symbiogenesis. But is this genes learning to manipulate bits – as they learnt to manipulate proteins – or the other way around? Certainly there has not been the time for genetic evolution to adapt to the control of digital technology, but given the speed that digital systems are evolving these systems could easily adapt to exploit their creators.

8. On Distributed Communications

The problems of modern telecommunications were encountered early in the history of the first optical data transmissions systems. Although only involving fires or lights on hills, the use of early forms of current protocols and the need to encode symbols was recognised. The cold war impetuous to the development of distributed communications networks is covered, as are the reliable digital communications systems utilising unreliable network repeater nodes developed by Paul Baran for the military packet-switching essential for distributed communications networks. Baran's outline for a distributed, multi-node, command and control system was published, without being patented, in 1964. Baran reasoned that "Not only would the US be safer with a survivable command and control system, the US would be even safer if the USSR also had a survivable command and control system as well!". Dyson draws out the historical connections of the cold war with the development of game theory and the analogy between telecommunication and nervous system control (e.g. reliable signalling in unreliable networks).

9. Theory of Games and Economic Behaviour

Here game theory is discussed in full. The application of recent developments in mathematics and logic to economics has been potent – seven Nobel prizes have been awarded to economists whose work was directly influenced by von Neumann and his formulations. Augmenting work on evolution and symbio-organisms, game theory points to the importance of co-operation and coalition for success in a large class of games.

Dyson also discusses parallels between nervous systems and economies, particularly their common reliance on statistical, rather than individually accurate signals. "Money is a medium for communicating value across distances and over time…the flow of money conveys and represents information". Economies are adaptive systems – economic plasticity allows the system to signal and then adapt to contradictions. The development of E-money realises one of the crucial conditions recognised by the cyberneticists for the development of intelligence: immediate feedback; E-money allows companies to do things and then immediately sense the results.

10. There's plenty of room at the top

Briefly noting Richard Feynmann's anticipation of nanotechnology, Dyson explains the benefits of being small; cheapness and efficiency – why there is 'room at the bottom'. But there's also room at the top, he claims, although life on a grand scale has been previously limited by gravity, chemistry and control problems. He goes on to give a history of the science of self-organisation, focusing on W. Ross Ashby whose work was later co-opted by the cybernetics movement. Ashby believed that any complex system (environment) will generate self-organising organisms. From there Dyson discusses top-down (Darwinian) organising principles compared to bottom-up (self) organising principles. "It is possible to construct self-preserving systems that grow, evolve and learn but do not reproduce, compete or face death in any given amount of time. It is also possible to view large, complex, systems, such as species, gene pools, and ecosystems, as information processing organisms providing a degree of guiding intelligence to component organisms whose evolution is otherwise characterised as blind." This quotation expresses a key theme of the book, that all informational systems are a kind of intelligence, but that intelligence may exist on a scale we find difficult to deal with. So evolutionary systems might have intelligence, although individual evolutionary events (e.g. the deaths of individuals or even individual species) do not show signs of intelligence. "Likewise, to conclude from the failure of individual machines to act intelligently that machines are not intelligent may represent a spectacular misunderstanding of the nature of intelligence among machines."

At this point Dyson sound dangerously group selectionist. He is indeed arguing for the existence of group level intelligence, however he is not a naïve group selectionist. If anything he is carrying the selfish gene argument to an extreme. Gene-level selection can operate at cross-purposes to individual-level selection - this is the theme of Dawkin's *The Extended Phenotype.* If we fully embrace this dissociation of the genes from the individual 'survival machines' that host them, we can begin to reconceptualise our theory of evolution to encompass genepools operating with some sort of collective 'purpose'. All Dyson's argument – or rather, his suggestion - requires is faith in some sort of endogenous organising factors to cohere the action of the genepool as a whole.

11. Last and First Men

This chapter contains the story of Olaf Stapledon and Lewis Fry Richardson. The pair, both pacifists, worked in the same ambulance crew in World War I France. I mentioned above Richardson's anticipation of cellular automata. As well as this, Dyson focuses on the fiction work of Stapledon, which dealt with ideas about the evolution of mankind. In the gloom of 1930s Stapledon wrote *First and Last men*, which traces the cosmic history of man from the perspective of a narrator at the end of time. In the book mankind had evolved a communal mind which allowed the overcoming of our lowly individual nature (which certainly appealed to the socialist principles of Stapledon). This communal mind was made possible by telepathy which resulted from symbiosis with a Martian invasion of a swarm composed of many 'sub-vital units'. Dyson applauds Stapledon's recognition that symbiosis is the next stage of human evolution, but claims that microprocessors are the telepathic units, and that this symbiosis is happening now, not 10,000 years in the future. Reading my summary you may be wary of such outlandish predictions, but Dyson's tone is measured and playful rather than hyperbolic and arrogant. This makes it more convincing than the hysterical claims of the post-humanist prophets. The simple fact is that machines have virtually unlimited memory and time. It may not be the beginning of global consciousness, but in Dyson's words, one thing is certain "global unconsciousness comes first". A group mind could be as ignorant of our functions as we are of neurons, and as oblivious to our passing. This is simply a logical extrapolation of the materialist/functionalist view of consciousness and a non-vitalist view of life.

It is strange that Dyson does not discuss the prospects for accelerated genetic evolution in the modern age. He hardly touches upon his own question in the introduction: "Do we remain one species or diverge into many?" Genetic science seems poised to create two new species, a genetically enhanced minority and a genetically impoverished majority (the 'GenRich' and the naturals as Lee Silver calls them in his book *Remaking Eden: Cloning and Beyond in a Brave New World*). Yet Dyson, although he is obviously comfortable with the idea that what it means to be human may change, neglects the prospects for genetic engineering in favour of discussion of human-computer symbiosis.

12. Fiddling while Rome burns

Both Samuel Butler and Olaf Stapledon saw that mind, once given a taste of time, would never rest until eternity lay within its grasp George Dyson.

Dyson's magical and masterful fugue is completed in the final chapter. He reiterates his conclusion that computers are massively increasing the speed of – mostly post-genetic - evolution and binding all organisms (all informational systems) into a single intelligence. He quotes Danny Hillis: "Memory locations are just wires turned sideways in time". Although religions, and science fiction in general, have always been preoccupied with the intervention of some higher, external intelligence, Dyson's assertion is that we have failed, or will fail, to recognise the super-ordinate intelligence that may have already developed among the global network of machines.

In conclusion I would say that Dyson's book will appeal to anyone who likes history and learning about the individuals who make it, or who likes to take in grand views and make connections between diverse fields of inquiry. The book closes on an almost spiritual note: "We have mapped, tamed and dismembered the physical wilderness of our earth. But, at the same time, we have created a digital wilderness whose evolution may embody a collective wisdom greater than our own. No digital universe can ever be completely mapped. We have traded one jungle for another, and in this direction lies not fear but hope. For our destiny and our sanity as human beings depend on our ability to serve a nature whose intelligence we can glimpse all around us, but never quite comprehend."