

Original Article

General Symbol Machines: The First Stage in the Evolution of Symbolic Communication

Thomas E. Dickins, School of Psychology, University of East London, United Kingdom. Email: t.dickins@uel.ac.uk.

Abstract: Humans uniquely form stimulus equivalence (SE) classes of abstract and unrelated stimuli, i.e. if taught to match A with B and B with C, they will spontaneously match B with A, and C with B, (the relation of symmetry), and A with C (transitivity). Other species do not do this. The SE ability is possibly the consequence of a specific selection event in the *Homo* lineage. SE is of interest because it appears to demonstrate a facility that is core to symbolic behavior. Linguistic symbols, for example, are arbitrarily and symmetrically related to their referent such that the term *banana* has no resemblance to bananas but when processed can be used to discriminate bananas. Equally when bananas are perceived the term *banana* is readily produced. This relation is arguably the defining mark of symbolic representation. In this paper I shall detail the SE phenomenon and argue that it is evidence for a cognitive device that I term a General Symbol Machine (GSM). The GSM not only sets the background condition for subsequent linguistic evolution but also for other symbolic behaviors such as mathematical reasoning. In so doing the GSM is not particularly domain-specific. The apparent domain-specificity of, for example, natural language is a consequence of other computational developments. This introduces complexity to evolutionary arguments about cognitive architecture.

Keywords: Symbols; stimulus equivalence; learning; modularity; domain-specific; canalization.

Introduction

This paper will present a conditional argument about the emergence of symbolic communication, and as such will constitute a hypothesis about a part of the evolution of language. Full, natural language is idiosyncratic to humans, for no other communication system exhibits the quality of recursion (Hauser,

Chomsky and Fitch, 2002) which is a property of syntax and undoubtedly the product of evolved cognitive machinery. However, one of the premises of the conditional argument to be presented is that the recursive property of syntax is dependent upon having something over which to operate – in this case, symbols. Symbols have distinctive properties that are not seen in other animal communication systems, and as a consequence, require an evolutionary explanation of their own. These properties will be described.

The other premises of the conditional argument are about the kinds of explanation we should be seeking when theorizing about the evolution of language. They might be termed epistemic premises or assumptions. As with the initial assumption that recursion has to operate over something, I am asking the reader to act as if these assumptions are the case, and instead to focus their critical effort upon the conclusions I seek to defend.

The first of these epistemic assumptions is a general one about cognitive science. Cognitive science assumes that there are computational processes operating within the brain that causally explain input-output relations in organisms. Much of cognitive science is about delivering functional descriptions of input-output relations, and trying to hypothesize the kind of algorithms that might deliver such regularity. What is more, cognitive science aims to reduce high level functioning to theories that rely only upon dumb, unthinking mechanisms. This paper is arguing about the characteristics of a dumb, unthinking mechanism that might underpin symbolic behavior.

The second epistemic assumption is about evolutionary theorizing specifically. Some contemporary brands of evolutionary psychology have argued from observations of domain-specific adaptive behavior for domain-specific cognitive mechanisms that are responsible for that behavior (see Dickins, in press; Samuels, 1998). This is sometimes referred to as a modularity commitment (see below). One concern with this approach is that although there are sound reasons not to believe that the brain is a totally domain-general processor it is not clear that behavioural evidence is sufficient to carve cognition at its joints. It is conceivable that one cognitive mechanism could instantiate a number of behaviors, and that one behavior could be the product of a number of mechanisms. The behavioural data will not always allow you to decide. Another concern is that for every novel mechanism hypothesized one is effectively hypothesizing a separate selection 'event'. All too easily, one could have a theory of the evolution of cognition that relied upon an unlikely number of fortuitous mutations. This paper sides with Hauser, Chomsky and Fitch (2002) in advocating a long, hard look at comparative evidence in order to be certain of behavioral discontinuities before advocating a novel cognition and attendant selection events. This paper maintains that symbolic behavior is just such a discontinuity, and will speculate about what can be said with regard to the cognition that enables it [1].

The aspect of language evolution to be discussed, then, is the emergence of

symbolic communication given the preceding assumptions. Communication will be defined, in line with the common view from behavioural ecology (see Hauser, 1996), as the transfer of information from an actor to a reactor, such that the behavior of the reactor is changed. This paper will argue that symbols convey a different order of information from more common signaling systems, and it is this that marks the discontinuity with other animal communication systems. The paper provides a discussion of what symbols are in terms of this difference and it will outline Deacon's (1997) hypothesis about the evolutionary transition to symbols within a communicative context to clarify this point.

Deacon is focused upon in some detail for his work embodies the premises just discussed. He grounds his hypothesis upon associative learning up until the point where symbols emerged, and as such Deacon adopts a parsimonious and comparative approach, albeit an abstract one. It is at the point where symbols emerge that his hypothesis will then be augmented with a discussion about the properties of a putative General Symbol Machine (GSM) that allows the formation of stimulus equivalence classes. This categorical ability is specific to humans and, it will be claimed, essential for symbolic behavior.

Information

Information is to be understood in terms of its role in reducing uncertainty. Through natural selection specific mechanisms will emerge that cause organisms to react to pertinent input. For example, a frog whose retina is stimulated by a fly crossing its visual field will produce an appropriate tongue-flick response that will lead to eating. The way in which the frog's visual system and tongue-flick system etc. are constituted renders the visual input *information* – the frog's systems can be in $1+n$ states and this input determines which of those states they will be in.

The manner in which information is 'transmitted' can be organized as follows: Cues convey information by being permanently on, or constantly present, for example the yellow and black stripes of a wasp. This is a continuous feature of a wasp's abdomen and indicates that the wasp carries a dangerous sting – fatal to some organisms, a painful irritant to others. This information reduces the uncertainty about whether or not to approach a wasp.

Cues require no more than perceptual salience and then a learnt association to be freshly established. The same is true of indices, or indexicals. The difference between indices and cues is that indices indicate the presence of something by dint of a causal relationship with that thing, such that smoke is the index of fire, foaming about the mouth is the index of scurvy.

Signals are unlike cues and indices. A signal gives information about the changing presence of something and as such can be on or off. Alarm calls are signals because they are only useful if on in the presence of danger and off in its absence. There is a sense in which signals are similar to indexical information for

they are a consequence of the thing they convey information about (Deacon, 1997; see below), but signals such as alarm calls are produced by an organism with a vested interest in the consequence of that signal being understood and acted upon. In other words, signals are used communicatively within a social group and as such are more effortful and have costs and benefits associated with their production and comprehension. For example, producing an alarm call makes one a target for the predator who is now aware of one's location, and acting on an alarm call opens one to possible deception. On the other hand, giving an alarm call can save your kin and acting on it can save your skin.

A symbol represents an object, event or state of affairs. Symbols are arbitrarily related to their referent, meaning that there is no natural relationship between a symbol and its referent. This arbitrary relationship is established and maintained through social convention. A symbol is also symmetrically related to its referent, such that the appropriate symbol can be produced in the presence of the referent and the appropriate referent can be produced or discriminated in the presence of the symbol. This key property of symmetry was first noted by Saussure (see Hurford, 1989, for a discussion).

The word <banana> is a symbol that refers to a certain kind of fruit. There is nothing in the term <banana> that would indicate its referent naturally; its use is entirely the consequence of the conventional linguistic history of English speakers [2]. When a banana is seen the word <banana> can be produced, and when the word <banana> is uttered the attention of the hearer is drawn to that kind of fruit. If a token of this kind of fruit is not present then the hearer will have activated an internal conceptual representation of a [banana], in this way reference can be displaced temporally and spatially (see Figure 1 overleaf).

The potential informational gains for organisms using symbols are great, for symbols allow the learning of others within a community to be transmitted and used by those without the direct experience. Simply by arranging symbols referents can be alluded to and novel situations involving those referents can be presented in their absence. In this way the reduction of uncertainty is spread beyond immediate domains.

Deacon (1997) has collapsed and refined the above taxonomy of information-bearing entities by proposing three main types – icons, indexicals and symbols – which owes much to the work of Charles Peirce, as Deacon makes clear. Icons are the significant addition to the above discussion, achieving informational content through bearing some similarity, for example, landscape paintings can be regarded as icons. Deacon's indexicals are the indices discussed above, however, he also includes signals within this kind due to their causal relationship with that which is signaled. His view of symbols is consonant with that already discussed. Deacon sees the transition from signals (indexicals) to symbols as the first major transition to language, as breaking the “symbolic threshold”. It is to this account that we now turn.

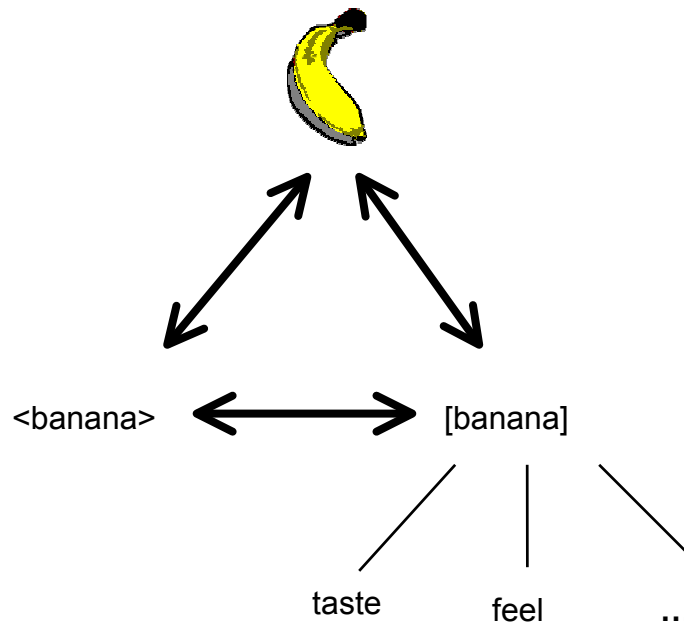


Figure 1: A symbolic relationship – the symbol <banana> is attached symmetrically to both the fruit and the concept [banana], which in turn is associated with a number of banana-related events and experiences such as taste, touch etc. At whichever point you access this categorical complex you can get to the other points – for example, on hearing the word <banana> you can accurately discriminate the fruit from other objects and this will also activate a conceptual schema.

Deacon's Symbolic Threshold

Under the definition of symbols adopted by this paper one could argue for a simple associative model for establishing symbolic reference. Our ancestors could simply have used novel vocalizations in the presence of certain objects and given enough stability and exposure an association would be formed between that vocalization and the object. A name would have been created. Deacon disagrees with such a pseudo-Skinnerian view, arguing that the correlation between symbols and their referents is not that frequent or strong in practice and as such, if symbols were merely associatively linked with their referents there is every chance that the relationship would quickly extinguish for most symbolic reference. What Deacon in fact believes is the somewhat counterintuitive claim that the 'correspondence between (symbols) and objects is a secondary relationship, subordinate to a web of associative relationships of a quite different

sort, which even allows us to reference impossible things' (1997, p. 70). It is in order to clarify this claim that Deacon introduces the tripartite taxonomy of icon, index and symbol.

The discussion of information sources could be used to endorse a passive view – an organism perceives an index and all of the information necessary to correctly orient the organism's behavior is provided by this index. It is just a question of downloading it. Deacon takes the opposite, behaviorally grounded view.

No particular objects are intrinsically icons, indices, or symbols. They are interpreted to be so, depending upon what is produced in response. In simple terms, the differences between iconic, indexical, and symbolic relationships derive from regarding things either with respect to their form, their correlations with other things, or their involvement in systems of conventional relationships. (Deacon, 1997, p. 71)

This view leads to the consequence that iconicity is not about brute similarity between the icon and the referent but is instead about the process 'based on recognizing a similarity' (Deacon, 1997, p. 71). As Deacon says, we can be very liberal about what features we construe as similar and make an iconic relationship out of practically anything. One can note similarities between a cheesecake and the moon, given enough inferential effort, but no one would claim a "natural" iconicity here. Likewise, a temporal or physical contiguity does not necessarily instantiate an indexical relationship, and conventional usage does not instantiate a symbol – it is only when we begin to use them as indexicals or symbols that they are such. An interpretive "decision" has to be made in each instance.

A consequence of this processing view is that we can begin to see the tripartite taxonomy as less defined. Icons, indexicals and symbols are not mutually exclusive categories and the same entity can potentially do the work of all three. Indeed, Deacon claims that these three classes of information are mutually interdependent to some extent. For example, we could imagine being in a foreign land and hearing a particular word used – <arnav> – on a number of occasions. As symbolic beings we might well realize that this is a symbol simply from the context in which the utterance is made but we would not have access to the conventions of what is, in fact, Hebrew linguistic culture and therefore we could not use the term symbolically. None the less, we might also note that this symbol is often used in the presence of certain creatures and learn that this is at least a likely index of the presence of rabbits. This guess will be heavily circumscribed by various assumptions about the level of categorization appropriate to the term but none the less, might covary sufficiently to facilitate some useful understanding. In this example we can refer to the information lost by not being part of the appropriate symbolic culture – if we spoke Hebrew and English we

would know that <arnav> meant the same as <rabbit> and not the more generic, and holophrastic <rabbits are present>, or <long-eared mammals are present> etc. So, the difference between icon, indexical and symbol is to be understood in terms of different levels of interpretation and these levels are hierarchically organized. In the <arnav> example, once symbolic understanding failed the strategy was to drop down to the next, indexical level and see what information we could use under the appropriate set of processes. Deacon gives the following example:

(As) human children become more competent and more experienced with written words, they gradually replace their iconic interpretations of these marks as just more writing with indexical interpretations supported by a recognition of certain regular correspondences to pictures and spoken sounds, and eventually use these as support for learning to interpret their symbolic meanings. (1997p. 74)

Deacon uses this idea as an intuition pump to drive the hypothesis that symbols are dependent upon indexical reference and indexical reference is dependent upon iconic reference. Could this hierarchical interdependence be the mark of an evolutionary transition to symbolic behavior and one based on simple learning behaviors?

Deacon discusses the different interpretative processes underlying iconic, indexical and symbolic representation. Iconic representation is the consequence of recognition, or of regarding the icon as like another thing. Sometimes this requires absolutely no processing effort at all, and Deacon uses the example of a bird scanning the bark of a tree to find a moth. The bird moves its head once – bark – twice – bark – thrice – bark, and so forth. As the moth's wings are very similarly patterned to the bark it gets missed. The bird would have to be looking harder for dissimilarities, rather than maintaining a process of similarity checking, to get fed – and there are always dissimilarities. To this extent the moth wings are iconic of the bark.

The obvious line to take when discussing the processes underlying indexical reference would be to argue for a learning history establishing links between foaming mouths and scurvy etc. However, as Deacon notes, many things can be said to have physical or temporal contiguity so there must be something more to this interpretative process. Deacon claims that it is critically dependent upon iconic skills, as we would expect given the preceding argument. He uses the example of smoke indicating fire:

The smell of smoke brings to mind past similar experiences (by iconically representing them). Each of these experiences comes to

mind because of their similarities to one another and to the present event. But what is more, many of these past experiences also share other similarities. On many of these occasions I also noticed something burning that was the source of the smoke, and in this way those experiences were icons of each other. (1997, p. 78)

So the extra process that is placed on top of iconic processing is that of noting repeated correlations, in this case between smoke and fire. The transition across the symbolic threshold is the next stage and this transition is, in Deacon's view, the establishment of relationships between indexicals, in a similar fashion to that in which indexicals are constituted by relationships between icons. In this way, symbols are not merely associatively linked to their referent. However, symbols do retain their indexical properties as a consequence of the inter-relationships between symbols as used in linguistic practice. Deacon exhorts us to think 'of the way a dictionary or thesaurus works. They each map one word onto other words. If this shared meaning breaks down between users ... the reference will also fail' (1997, p. 82). The intensionality of a linguistic symbol or word is established and maintained by the word-word relations, whilst the indexical element or word-use provides the extension – word-word relations 'allow words to be about indexical relationships' (Deacon, 1997, p. 83). Indeed, contextual information provided by words often supports our comprehension of new terms.

How could this symbolic system establish itself in an ancestral population? Deacon's claim is that what establishes a symbol-symbol relation is a form of insight learning. He supports the notion by discussing child development and lays claim to bursts of learning within the language domain that are indicative of ongoing insights. It is at this point in Deacon's theory that there is a gap to be filled.

Inferential Effort

The standard view of language acquisition in modern human infants is that much of it is governed by innate mechanisms. For syntactic elements of language these mechanisms are highly structured modular devices that effectively impose a set of principles on a child's learning of their native tongue. For word learning – i.e. basic symbol acquisition – there is less evidence of a specific device, and instead much discussion about canalizing learning with a number of innate constraints, such as a whole-object bias, sensitivity to ostensive cues, novel objects and novel speech sounds (Bloom and Markson, 1998). These constraints triangulate the referent to which a given word is related such that an infant hears the novel sound being uttered by an adult, looks to ascertain the direction of attention (primarily from gaze direction), fixes the new object and assumes the word refers to that whole object.

It would appear that word learning does rely upon associative learning, but this learning is heavily directed in order to deliver specific associations. Deacon would argue that initially words are acquired as indices and only later do they gain intensional properties once symbol-symbol relations are established. What is surprising is the speed with which children acquire words and the lack of explicit associative training that they receive, specifically negative training. To some extent the canalization argument will account for this effect, for it reduces the number of possible associations that can be made to “sound x ” goes with “whole object y ”, but under normal associative learning paradigms one might expect a few trials to be undertaken before such a novel link is made (see below).

It would appear reasonable to look at the kinds of associative learning that might be operating under the canalizing constraints. Some discussion of the nature of the learning might actually enable us to say something more about the inferential effort required to interpret something as a symbol.

Using a straightforward matching-to-sample (MTS) paradigm with abstract stimuli Sidman (1971, 1986, and 1994) was able to demonstrate a number of emergent relational properties in human participants. A simple MTS procedure consists of a training phase and then a test phase. In the training phase participants are taught, through feedback, to pair abstract and unrelated stimuli according to an undisclosed pattern. The experimenter might have three sets (A, B and C) each of three stimuli (1, 2, and 3; see Figure 2a) and would train A1-B1 (which means that in the presence of sample stimulus A1 the comparison stimulus B1 should be selected from B1, B2, and B3), and A2-B2 and A3-B3; and then B1-C1, B2-C2, and B3-C3.

Figure 2a: Nine abstract stimuli for use in the formation of three three-member stimulus equivalence classes. The classes to be formed are A1-B1-C1, A2-B2-C2, and A3-B3-C3. In this example characters from the Klingon alphabet have been used. None of the characters have a natural relationship within their categories.

	A	B	C
1	E	Ƨ	Ƨ
2	I	Ƨ	Ƨ
3	Ƨ	Ƨ	Ƨ

Figure 2b: A test of A1-A1 identity (with outlined correct response)

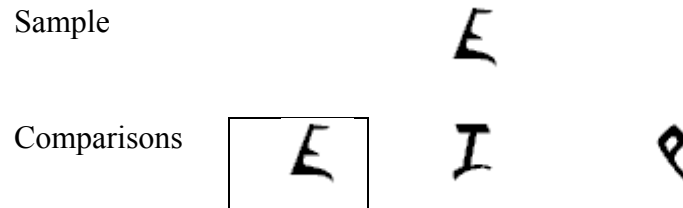


Figure 2c: A test of B1-A1 symmetry (with outlined correct response)



Figure 2d: A test of A1-C1 transitivity (with outlined correct response)

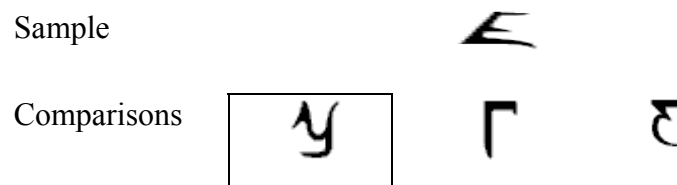


Figure 2e: A test of full C1-A1 equivalence (with outlined correct response)

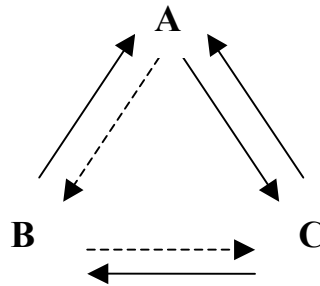


In the test phase participants receive no feedback and are tested on a number of trained and untrained relations between the stimuli in a random order. The key tested but untrained relations are:

1. **Identity** (Figure 2b) – for example, A1 is presented as a sample and the participant must choose A1 from A1, A2, and A3 (A1-A1);
2. **Symmetry** (Figure 2c) – for example, B1 is presented as a sample and the participant has to choose A1 from A1, A2, and A3 (B1-A1) thereby reversing the trained relation;
3. **Transitivity** (Figure 2d) – for example, A1 is presented as a sample and the participant has to choose C1 from C1, C2, and C3 (A1-C1) thereby combining two trained relations at their common node, in this case B1;
4. **The equivalence relation** (Figure 2e) – for example, C1 is presented as a sample and the participant has to choose A1 from A1, A2, and A3 (A1-C1) thereby combining two trained relations at their common node, in this case B1, and reversing them.

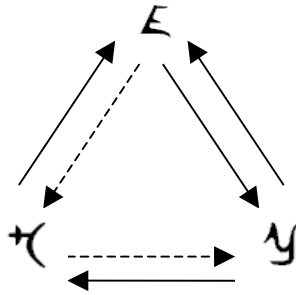
When a participant is able to produce all of these untrained relations from the set of stimuli they are said to be in possession of a mathematical equivalence set, sometimes referred to as a stimulus equivalence class (Figure 3a and b).

Figure 3a: The stimulus equivalence class paradigm. The trained relations are represented by the dotted lines. All other relations are emergent, with no training.



This phenomenon is of great interest because the emergent relations are completely untrained and occur after some very simple associative learning. Of particular interest is the fact that this is done using arbitrary and unrelated stimuli, which is a property of a symbol with relation to its referent, and this link is turned into a symmetrical one apparently spontaneously, which is also a symbol property.

Figure 3b: A stimulus equivalence class of Klingon characters, as trained in Figure 2. The trained relations are represented by the dotted lines. All other relations are emergent, with no training. Compare with Figure 1.



The fact that most vertebrate species can acquire the basic trained relations, but, with one or two possible exceptions (Schusterman and Kastak 1993), humans are the only species that can acquire the full set of equivalence relations is a tantalizing hint that this ability might be linked to language. Furthermore, pre-linguistic infants and those without a linguistic capacity appear unable to form equivalence classes (Barnes et al., 1990; Devany et al., 1986) although some researchers argue against this (Carr et al., 2000). Finally, D. W. Dickins et al. (2001) conducted an fMRI study on stimulus equivalence class formation that showed significant activity in the dorsolateral prefrontal cortex and activity in parietal and other regions. The speed and accuracy of response on tests of trained relations and transitivity was found to be positively associated with the degree of left lateralization across the group of participants. There was no Broca's activation whatsoever, despite the use of readily nameable stimuli in this experiment which suggests that no sub-vocalization was occurring during the tasks. Sub-vocalization of this sort is something that Horne and Lowe (1996) have attributed to the ontogenetic causes of stimulus equivalence class formation, which they see as a consequence of language (see below). Whilst not conclusive evidence for this ability coming under separate control from the language faculty as a whole it certainly suggests a more general order of cognitive functioning for stimulus equivalence processes.

If symbols are part of a stimulus equivalence class what is the exact nature of the relation? Infants learn a sound-to-object relation, and not an object-to-sound relation. The ability to use this symmetrically is a part of the full equivalence relation. It is also likely that infants already have a notion of the indices of the named object. These indices will also be treated symmetrically (although they

may be grouped under a concept – see Figure 1). Place (1995) has argued that an equivalence class consisting of a name, an object and at least one index of that object is the core of a conceptual representation of that object. Deacon would add that once the name in a particular equivalence class is associated with the names for others then a set of higher-order classes could be formed. Thus hearing a name would stimulate a vast set of associations between classes and conjure up many possible extensions. So, stimulus equivalence class formation potentially provides symmetry, the core symbolic property, and later on symbol-symbol relations.

Stimulus equivalence classes are categories. A vocalization within an equivalence class will evoke all the other aspects of that category. It would appear that we assume equivalence class categories under the very specific conditions invoked by an MTS schedule, but stimulus equivalence class formation is potentially the product of a cognitive mechanism with fairly indiscriminate application. This mechanism is delivering relations between previously unrelated stimuli, of any kind, and as such cannot be relying on straightforward associative learning – even though it is predicated upon early trained associations between some stimuli. This mechanism could be described as doing inferential work, for it allows the organism to go beyond the data presented. As we have already seen, Deacon believes that the kind of information one infers from a situation is dependent upon the interpretative effort one makes. The mechanism underpinning stimulus equivalence class formation could afford the effort required to break through the symbolic threshold.

Evolved mechanisms and a General Symbol Machine

The standard view from evolutionary cognitive science is that natural selection has led to a modular cognitive architecture (Cummins et al, 2003; Dickins, in press; Tooby and Cosmides, 1992). Modules are to be thought of as domain-specific packages of computations and/or representations, the domains being set by ancestral adaptive problems (Samuels, 1998). This means that some modules are highly content based, others highly process based and some are both process and content based. The obvious question to ask is whether we can conceive of the mechanism delivering stimulus equivalence class formation as a symbol-producing module. We might call this a General Symbol Machine (GSM) given its ability to form relations between any kind of stimuli.

Any content of a module that is established through natural selection has a truth-value such that it matches a real contingent problem in the world at the time of selection. The putative GSM is odd because it is a module that enables symbols to be formed and as such has no content, merely algorithms, or processes. Symbols have the adaptive advantages suggested above, but there is no sense in which symbols were out there in the world to provide a selection

pressure and in turn create a module with internal and veridical representations of the symbols in the world. Instead, as we are discussing the selection of a process, the GSM could only be a form of computational module. This means that a GSM is an inference machine of sorts, a hypothesis imposition device that looks for symbolic (equivalence) relations that might be of use in order to further reduce prior uncertainty. Thus under this model, a GSM running on stimulus equivalence principles sets up the symbolic relation as a hypothesis and this is subsequently tested.

The successful attachment of a vocalization to an object, and consequently to the indices of that object, establishes an equivalence class that is empirically demonstrable. The vocal attachment is entirely arbitrary in evolutionary terms, but in ontogenetic terms it is not for it is an established fact of the linguistic community and there to be learnt. Nonetheless, once the attachment is made categorical work can be done. But the mechanism that enables this, a GSM, simply forms equivalence classes between any stimuli that are presented in a specific way. As such it assumes the existence of a kind of category. The likelihood of this assumption being instrumental must have been high in ancestral times, hence the selection of the process.

Any symbol relations established by a GSM might be of use to other processing systems. An entailment of the GSM hypothesis is that the evolutionary emergence of a GSM is a precursor to any other higher-order symbol processing. A GSM will provide a rudimentary symbol that would have to be later refined by other linguistic mechanisms (predominantly syntactic) in order to be incorporated into a full natural language – for example, words have characteristics extra to the core symbol properties outlined, such as morphology. Equally, the symbols of a GSM could be fed into mathematical processes. This is not a simplistic statement of the order “first came symbols, then came language” for it is undoubtedly the case that the onset of a functioning GSM would have effects upon extant cognitive systems and also would shape the possible selection future of hominid cognition. A GSM would be more than a catalyst for it would alter the cognitive economy of hominids and the selection pressures they faced. As such a GSM would construct a particular cognitive niche for our ancestors (cf. Laland et al, 2000).

Past discussions of modularity have focused upon domains such as social reasoning and the detection of cheats (Cosmides, 1989). Domains of this order are very specific, dealing with a clear set of costs and benefits and a defined class of inputs. A GSM is *not* of this order for it allows the indiscriminate formation of an equivalence class between any set of stimuli, but the nature of this process is clearly prescribed and as such cannot be regarded as a totally general learning mechanism. But despite these prescriptions stimulus equivalence classes only emerge in the lab after exposure to many trained relation trials on an MTS paradigm. This very fact might lead one to question the notion of a GSM

underpinning this learning – for one might expect such a mechanism to be fast especially if it underpins word acquisition in infants. However, as described above, word acquisition is in fact the consequence of some specific canalizing constraints triangulating learning. This learning effectively becomes a focused MTS procedure with explicit guidance as to which choices to make and one might expect this to speed up acquisition relative to our general assumptions about associative learning. What this description of canalization does not provide is an explanation of how the symbolic relation of symmetry is formed, but the addition of a GSM hypothesis does.

Within the stimulus equivalence literature there is debate about whether language is an ontogenetic precursor to the ability to form such classes (Horne and Lowe, 1996) or vice versa (Dickins and Dickins, 2001). The language-first hypothesis rests on a Skinnerian view of language acquisition that fails to account for rapid acquisition. However, Place (1995) refined this and suggested that language indeed leads to the stimulus equivalence ability in infants, but that once that ability is in place it speeds up later word acquisition. Despite this innovation Place's hypothesis fails to account for the establishment of symmetry in the first instance, arguing instead that all of the equivalence relations are somehow transferred from a Skinnerian training of initial words. There is no evidence whatsoever to support Skinnerian learning of words in infants. Under the GSM hypothesis stimulus equivalence has ontogenetic primacy and facilitates word acquisition. When stimulus equivalence classes are formed in the lab it is a lengthy procedure but here abstract stimuli are being used without any canalization – the task confronting the participant is as difficult as that confronting an infant with no innate constraining biases [3]. Furthermore, in the language situation an infant is busy forming full equivalence relations between a vocalization, an object, and indices of that object as well as other learnt symbols. This will greatly facilitate subsequent learning, as Deacon's dictionary example suggested.

To summarize, the putative GSM, as a mechanism that *solely* delivers stimulus equivalence classes, is not content based, it is a process, and it is not domain-specific because we would appear to be able to form equivalence classes with any stimuli. Given this the GSM is not a module. However, the generality of stimulus equivalence class formation is a phenomenon noticed in the laboratory. If the relation between the ability to form stimulus equivalence classes and symbolic behavior is as suggested, then the addition of canalizing constraints completes the GSM, making it domain-specific. One might wish to refer to the GSM as an indirectly exapted learning mechanism, such that the onset of other cognitive abilities augmented a fairly general learning ability (Dickins and Dickins, 2002).

Conclusion

The evolutionary story might run as follows. At some point in our ancestry we developed the ability to form stimulus equivalence classes. This allowed primitive symbolic relations to be formed, which enabled a greater range of inferences to be made about categorical properties of the world and also set the stage for subsequent evolutionary changes leading to linguistic, mathematical and other symbol processing. In the case of language, the learning situation for infants was further augmented by the onset of key canalizing biases that made the training phase for linguistic equivalence classes quicker and more accurate. Some of these biases, such as the whole object bias, may have been in place already as a key categorical skill, but it is the completion of the full suite that led to the behaviors we see in modern humans.

The GSM hypothesis that has been tentatively proposed here argues that the ability to form stimulus equivalence classes is a species-specific cognitive ability produced by a computational device with a delimited but reasonably general domain of operation. The domain is simply that of any stimuli but the delimiting factor on the process is the presentation of those stimuli. In the case of word acquisition this task is made simpler and more constrained by the existence of well-documented perceptual biases. A GSM imposes inferences about category membership on the world and it allows arbitrary vocalizations to be incorporated into such categories such that they can do symbolic work. What remains to be explained is how and why such arbitrary vocalizations were ever produced, but for now we have a candidate mechanism that, with further study, might provide an explanation of the rudiments of symbolic relations.

Received 17th June, 2003, Revision received 18th August, 2003, Accepted 22nd September, 2003.

Notes

1. Readers might note that the definition of symbols given later in the paper conforms to the symbolic feats of some chimpanzees (Savage-Rumbaugh et al., 1993). However, these feats are accomplished after many, many trials which arguably contain hidden operants (Sundberg, 1996). Even if no operants are present, *contra* Sundberg, these accomplishments, although impressive, are unlike the symbolic competences of human infants that acquire symbols fast and with no training. Infant acquisition is discussed later in the paper.
2. It is worth noting that there is nothing in the alarm calls of, for example, vervet monkeys that would indicate a given predator naturally. Arbitrariness *per se* is not a key symbolic property – symmetry is. None the less, the form of

arbitrariness associated with alarm calls is different from the kind established through social convention, for it has an innate component.

3. This is the same order of task confronting various chimpanzees – see footnote 1.

References

- Barnes, D., McCullagh, P. D. and Keenan, M. (1990). Equivalence class formation in non-hearing-impaired children and hearing-impaired children. *The Analysis of Verbal Behavior*, 8: 19-30.
- Bloom, P. and Markson, L. (1998) Capacities underlying word learning. *Trends in Cognitive Science*, 2 (2): 67-73.
- Carr D, Wilkinson K. M., Blackman D. E., and McIlvane, W. J. (2000). Equivalence classes in individuals with minimal verbal repertoires. *Journal of the Experimental Analysis of Behavior*, 74: 101-114.
- Cosmides, L. (1989) The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31: 187-276.
- Cummins, D., Cummins, R. and Poirier, P. (2003) Cognitive evolutionary psychology without representational nativism. *Journal of Experimental and Theoretical Artificial Intelligence*, 15: 143-159.
- Deacon, T. W. (1997) *The Symbolic Species: The Co-evolution of Language and the Human Brain*. Allen Lane: The Penguin Press.
- Devany, J. M., Hayes, S. C., and Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior*, 46: 243-257.
- Dickins, D. W., Singh, K. D., Roberts, N., Burns, P., Downes, J. J., Jimmieson, P., and Bentall, R. P. (2001). An fMRI study of stimulus equivalence. *Neuroreport*, 12(2): 405-411.
- Dickins, T. E. (In press) What can evolutionary psychology tell us about cognitive architecture? *History and Philosophy of Psychology*.
- Dickins, T. E. and Dickins, D. W. (2002) Is empirical imagination a constraint on adaptationist theory construction? *Behavioral and Brain Sciences*, 25 (4): 515-516
- Dickins, T. E. and Dickins, D. W. (2001) Symbols, Stimulus Equivalence and the Origins of Language. *Behavior and Philosophy*, 29: 221 – 244.
- Hauser, M.D. (1996) *The Evolution of Communication*. Cambridge MA: MIT Press.
- Hauser, M.D., Chomsky, N. and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298: 1569-1579
- Horne, P. J., and Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65:

183-353.

- Hurford, J. R. (1989) Biological evolution of the Saussurean Sign as a component of the language acquisition device. *Lingua*, 77: 187-222.
- Laland, K. L., Odling-Smee, J. and Feldman, M. W. (2000) Niche construction, biological evolution and cultural change. *Behavioral and Brain Sciences*, 23: 131-175.
- Place, U. T. (1995) Symbolic processes and stimulus equivalence. *Behavior and Philosophy*, 23/4: 13-30.
- Samuels, R. (1998) Evolutionary psychology and the massive modularity hypothesis *British Journal of the Philosophy of Science*, 49: 575-602
- Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L. and Rumbaugh, D. M. (1993) Language Comprehension in Ape & Child. *Monographs for the Society for Research in Child Development*, 58: 1 - 256.
- Schusterman, R. J. and Kastak, D. (1993) A California Sea-Lion (*Zalophos californianus*) is capable of forming equivalence relations. *The Psychological Record*, 43: 823-839
- Sidman, M. (1971). Reading and audio-visual equivalences. *Journal of Speech and Hearing Research*, 14: 5-13.
- Sidman, M. (1986). Functional analysis of emergent stimulus classes. In T. Thompson and M. D. Zeiler (Eds.), *Analysis and integration of behavioral units*. Hillsdale, N.J.: Erlbaum.
- Sidman, M. (1994). *Equivalence Relations and Behavior: a Research Story*. Boston: Authors' Cooperative inc.
- Sundberg, M. L. (1996) Toward granting linguistic competence to apes: A review of Savage-Rumbaugh et al's Language Comprehension in Ape & Child. *Journal of the Experimental Analysis of Behaviour*, 65 (2): 477 - 492.
- Tooby, J. and Cosmides, L. (1992) The psychological foundations of culture. In: Barkow, J H, Cosmides L and Tooby J (eds) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press.