

EXTENDIBILITY OF ACTIVITIES AND THE DESIGN OF THE NERVOUS SYSTEM

Brian D. Josephson

Department of Physics, University of Cambridge, Cambridge CB3 0HE, UK

Abstract

The paper presents an integrated account of the workings of the brain, built up using ideas from Minsky's *Society of Mind* approach, computer science, and developmental and evolutionary psychology. In the course of development a collection of agents is trained by a virtual 'master plan' that has, as the outcome of evolution by natural selection, acquired the capacity to bring about their cooperative and harmonious development. Evolution has discovered both powerful computational devices such as threads and classes of software objects, and major ways of ensuring fitness such as communication by means of natural language. These two aspects work together in that the one kind of discovery is needed for the other to be possible. The possibilities of the approach are explored through examples such as walking, fetching objects, planning and language use.

Keywords

Nervous system, agents, society of mind, development, evolutionary psychology, language.

Introduction: the variety of approaches to explaining mental functions

The central problem of the neurosciences is to explain why the particular structure possessed by the nervous system gives rise to the particular behaviour that is observed. A great range of approaches to this problem, each with its own capabilities and limitations, has been developed. Direct observation of the brain combined with correlating the results of these observations with observed behaviour reveals many mechanisms underlying behaviour, but explanations that result are of a qualitative character only. Neural network models (e.g. Elman et al. 1996, Churchland 1995) can explain how skills are acquired, but normally deal with specific contexts only and do not address themselves to the more formidable complexities of real situations. Again, Minsky's *Society of Mind* (1987) pictures the mind as a collection of interacting agents. Many aspects of behaviour can be discussed within this paradigm, but it has little in the way of explanation of how the whole collection is organised. Classical artificial intelligence attempted to explain complex skills in terms of algorithms, but there are a number of situations for which an algorithmic approach is inapplicable. The approach of artificial life avoids this limitation, but as yet models low-level cognitive processes only.

Other approaches start from the perspective of

development. The psychologist Piaget investigated general principles behind development that complement the more usual studies of specific aspects of development, and Josephson and Hauser (1981) working within a Piagetian framework discussed the logic associated with specific increments of the developmental process fitting together into an integrated whole, an idea subsequently applied to the development of language by Josephson and Blair (1982). The hyperstructure approach of Baas (1994), and the Evolutive Systems approach of Ehresmann and Vanbremeersch (1987) are formal approaches that attempt to describe the way structure builds up and self-organises during development. These approaches, unlike the ones discussed in the previous paragraph, are limited by being primarily of a descriptive character.

Finally, two approaches from the viewpoint of evolution are those of evolutionary psychology (Pinker 1997), and genetic algorithms (Back 1996), the former being more conceptual and the latter more computational. Into this class can be added the robotics models of Brooks (1986), designed from an evolutionary point of view.

Basic principles of a synthesis: agents as building blocks of the mind

All of the above ways of tackling the problem of explaining the brain are unsatisfactory in that they concern themselves with only aspects of the problem and ignore the rest. Here a synthesis of approaches is proposed, which by providing a more all-embracing point of view may be able to go beyond such limitations. This synthesis invokes mainly the concepts of the *Society of Mind*, *neural networks*, and

evolutionary psychology. The basic logic is while that individual neural networks appear to be limited in the sophistication of what they are able to achieve, they are suitable nevertheless to act as the *agents* of the society of mind model, which is capable of more complex behaviour.

Neural networks are hypothesised to develop so as to become effective agents by learning to fulfil specific *requests*. For example, the process of walking demands certain balancing abilities, defined as an act of deploying muscles in such a way as to prevent falling over. This activity is supported by sensory systems that provide information related to falling over or the likelihood of falling over. It is assumed that a particular neural system generates the balancing response taking into account such specific sensory information and any other relevant information. There is a default (initial) response that enables the system to get started. When a certain condition is satisfied, e.g. the child has pulled herself to a standing position, a signal constituting a 'balancing request' is generated and the system attempts to answer the request, starting with the default response and then using feedback from the outcome to modify weights in the system that generates the response. Thus the balancing agent is a neural system (that has come into existence as a result of natural selection, as will be discussed) that not only has appropriate inputs and outputs but also has the right weight-adjustment algorithms for learning this particular task.

A similar discussion can be given for other kinds of agent involved with walking, such as that which orients the body to face an object it wishes to walk towards. In the following, a number of examples of various levels of sophistication will be discussed, leading to the general conclusion that the concept of systems having evolved to train various categories of agents is very widely applicable.

Agents and computation

An assumption will be made now that is implicit in the society of mind idea, namely that many aspects of neural functioning are isomorphic to the workings of computer programs executing similar tasks (which is not the same as saying that the brain as a whole is a computer; it is better considered as a complex mechanism that has embedded in it a number of computer-like subsystems). This applies not only to the way activities are called in particular sequences and branch according to specific criteria, but also to processes such as assignment of values to variables and parameter passing.

Assigning values can be done in more than one way. In cases such as learning how hard to press the foot on a pedal, it is likely that adjustment of weights involving neurons specific to the task is the key process. In other situations the value consists of an identifier or another variable, which we take to be

associated with its own agent. If, for example, we learn the number of a house, the number is almost certainly not represented by a signal whose strength represents the number. More plausibly, the number has an agent associated with it (and when a new number is learnt an agent is created for that number), and when we learn the number of the house a link is made between an agent for the house and the agent for the number. The link needs also be such as to carry the implication that the relationship between the two constructs is one of number (rather than, say, one of colour), and so it needs to include the equivalent of a switch which switches on if an agent for 'number of' is active.

Assignment and parameter passing demand special kind of neural hardware, allowing the hardware corresponding to particular agents to be manipulated in very specific ways. We assume that nature has evolved such mechanisms as are needed. Such mechanisms do things like activate particular agents for the equivalent of read and write operations.

The problem of organisation of agents

A question not very well addressed by Minsky's analysis of the society of mind idea is how it is that the agents are able to act in an integrated, comprehensible manner. In biology generally this is seen as a matter of design resulting from natural selection. There are many kinds of structures that might have been formed from strands of DNA different from the ones actually found but, owing to natural selection, generally speaking it is only ones that perform some biologically useful function in an efficient manner when they interact with each other in the right way that are found in nature. In the same way, only agents in the brain that in some sense interact together in an efficient manner are found in the brain. As will be discussed in more detail later, evolution discovers what kinds of agents are useful and how to connect them together so that they generate the best outcome.

This statement must be qualified in that it is only in the *developed* brain that efficient cooperation is found. The correct statement then is that the structures are designed to foster the development of cooperation during the developmental process. Processes must exist that cause efficient cooperation to develop in most cases. In other words, there is an implicit master plan, leading through various intermediate stages which, by a process of the form of putting requests to agents that they have to learn how to answer, lead to a developed state where the agents end up cooperating in an effective way.

Exceptions and the throw-catch mechanism

In artificial intelligence programs an attempt is made

to specify completely the algorithms that are used. Unfortunately this goal of complete specification appears unachievable in practice. The approach here suggests an alternative. The details are not innately specified; instead we have a system which is specified in outline in terms of a number of basic ways of dealing with problems, and such solutions as are found are used to build up the agent system which therefore reflects what has been learnt from experience as well as innate knowledge. A powerful scheme for building up complex systems is the so-called 'throw-catch mechanism' of object-oriented programming (Niemeyer and Peck 1997, section 4.5). According to this, when a process encounters an error, it deals with it by saving details of the error and transferring control to a new process specified as being able to 'catch' the error (called an exception in this context). This mechanism, which operates in the background, means that the sequence of operations does not have to be specified explicitly in all cases; events determine what the sequence will be. In the context of the mind, a typical exception might be 'about to lose balance' or 'in the wrong place'. Thus whatever one is doing that causes one to lose balance, a balancing agent will be called upon to restore balance. If ever something happens that fails because one is not in the place where one has to be for the current action to succeed, an approach agent is called upon to try to deal with the exceptional circumstance.

The 'exception' concept is applicable equally to the situation of trying something new. In this case the exception consists in the fact of having become competent at a particular process, and the 'catching'

construct consists of a more advanced process that can take advantage of what has now been learnt.

Emergence and evolution

Let us now examine the evolutionary aspects in more detail. Evolution is a process in which gradual changes occur (in the present context, changes in the nervous system), associated with increased competence at various forms of activity and thereby increased fitness. Some changes involve increased competence at particular forms of activity, and others the emergence of new forms of activity. In the picture discussed here these changes are associated with changes in agents and schemes by which the agents develop. Frequently these changes are really changes in emphasis: a process occurs initially on occasion as a chance outcome of mechanisms that already exist, and if this process contributes to overall fitness then modifications to the nervous system that facilitate such a process may contribute sufficiently to fitness that there is a drift of the genetic population in that direction.

Most changes have detrimental effects in an evolved system and only coordinated changes, taking into account the specifics of the behaviour concerned, are likely to be of benefit. Thus new directions in evolution are likely to be the outcome of some accidental coordinated change which is of overall benefit. Once a mutation has occurred, further changes can occur enhancing that particular benefit, or going off in a different direction, taking the initially changed system as a starting point (see fig. 1).

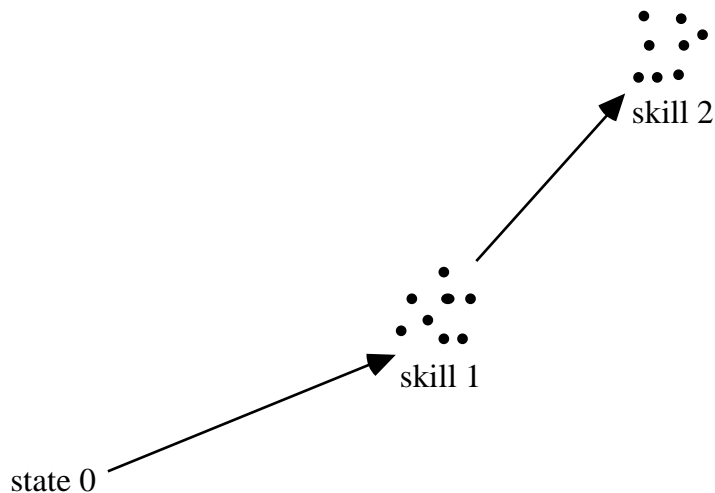


Figure 1. The diagram depicts a nominal 'space of skills' or space of operations that may be acquired during development. The dots represent valuable skills such as balancing or walking that are clustered in particular zones in the overall skill space. A training system explores particular zones in the skill space, and matching the training process to the distribution of valuable skills in the skill space enhances the extent of acquisition of the relevant skill. Natural selection discovers valuable training processes, symbolised by the two arrows in this diagram.

Evolution can solve the problems of designing systems that can develop functioning agents in a

sequence of steps rather than all at once, as illustrated practically in the approach to designing robots of

Brooks (1986). Each incremental step demands specialised agents that can develop using appropriate training procedures, and gradually evolution discovers procedures that perform this task efficiently. It must also develop systems that organise the 'master plan' referred to above, which impels the system gradually, through tasks of increasing difficulty involving new kinds of agents, progressively being trained to act in the way dictated by the master plan.

Evolution cannot know in advance what the effect of a modification in design will be, but if some change produces an effect such as making it more likely to experiment with taking steps with an outcome beneficial to the individual, then evolution can shift in that direction. It can be said that evolution explores various ways of combining existing abilities and occasionally stumbles, in a rudimentary form, upon a combination corresponding to a major advance. This rudimentary form then evolves to more effective versions of the same latent capacity.

These ideas are illustrated graphically in fig. 1, which may be thought of as depicting a space of available operations, points in this space which are of value according to some appropriate criterion being indicated by dots. From the original state 0 there is a direction one can go to a zone 1, high in good options. A suitable mutation leads to nervous system modifications leading one to explore this zone and discover good possibilities. From here a different mutation can lead one to explore more advanced options still in zone 2.

The concept so far

At this stage we pause to assemble together the various strands of the argument. One is that agents (which may have specific neural correlates) can be used to model various aspects of computation, including specific algorithms that might be used in an artificial intelligence simulation. The model goes beyond artificial intelligence simulations in that the agents are supposed to develop by experience, so that a perfect system does not need to be present at the beginning. New agents can be created during development so as to cope with new situations, though the classes of agents may be largely specified innately in a way that reflects discoveries that were made about various types of situations during the course of evolution. The developmental process is highly organised, taking into account the specific characteristics of all the various kinds of agents. A 'master plan' has evolved to make this process occur in an orderly manner such as we see in nature that approximately optimises the fitness of the developed organism. The brain is a very complicated system but it is complicated out of history and out of necessity. The complexity reflects the wide range of types of thing that the brain does and all the modifications that have been added during the course of evolution to cope with imperfections of existing

systems, and is similar in nature to the complexity of man-made machines.

The above discussion has been conceptual, but it is not the intention that the analysis remain confined to a speculative or philosophical level. It constitutes a system of ideas that can in principle be integrated closely with both experimental observation and computer modelling. Observation of how children actually develop skills should allow pinning down of the specific agents and their means of development. Such models can be integrated with neurological research to try to identify the neuronal systems corresponding to the various kinds of agents. The evolutionary viewpoint is valuable in indicating an approach starting with primitive systems and moving on to more advanced ones regarded as extensions to the simpler ones. Computer simulation can be used to study the models in more detail. The proposals are thus potentially much more concrete than were the original ideas of Minsky. The process described has the character of solving a jigsaw puzzle by trying various pieces and seeing which ones fit best.

The remainder of the paper will remain theoretical, but it will be less abstract in that it will concern itself with concrete examples, starting with the basic phenomenon of walking and going on, through more complex forms of behaviour, to language. The analysis of the latter in part answers the question of whether a system working on the basis of the principles described can really give rise to the complexities and subtleties of real behaviour.

Analysing walking

The first step of the analysis of walking is to consider various stages of walking, e.g. rising to a vertical position, standing, talking a step, joining steps together into a walk, changing direction, walking to an object, going round obstacles, etc., all of which have advantages in themselves from an evolutionary point of view (e.g. rising being associated with an improvement in the visibility of the surroundings). Thus evolution first develops systems that can acquire the ability to rise, then systems possessing the ability to stand, and so on. One then assumes that there is an agent associated with each of these operations, i.e. a rising agent, an agent for balance while standing, a stepping agent, each agent being a system which when activated appropriately performs the corresponding action. The stepping agent, to take an example, both takes a step with one foot and biases the balancing process so that a falling forward process takes place. The training process uses trial and error to coordinate these two aspects of stepping.

Individual agents perform relatively simple components of an action (balancing, taking a step, etc.), and so can be trained by relatively simple procedures. As is implicit in the work of Josephson and Hauser (1981), Brooks (1986) and Baas (1994), an

accumulation of simple advances can lead to a complex skill, as illustrated by the example of the sequence indicated above leading ultimately to controlled movement through an environment.

Continuing the analysis, similar considerations to stepping apply to other components of walking and its use. The ability to move short distances by walking confers some advantages, but the system can be enhanced still more if it is possible to direct walking to some specific destination. This demands some ability to turn towards an object, which involves specifically linking together a system for turning (for which an evolutionary pressure is the advantage of being able to turn to look at something outside the field of vision) with the visual system, and then training the system to produce the correct outputs by trial and error. We have once again the theme of a specific system conferring benefits, which can evolve through a crude system to a more refined system.

Other kinds of agents may learn to handle situations such as dealing with obstacles, at first autonomously and later in coordination with getting to a goal. A typical system design would involve modules that initially learnt to perform particular ways of dealing with obstacles as an isolated activity. At a later stage, learning to activate the module most effective in the context of achieving a specific goal would take place. More advanced modules could handle activities such as planning a route to reach a destination, which involve more complex considerations such as modules specifically connected so as to be able to learn a sequence, which could be involved for example in planning a route. These will not be discussed here, but are of the same general kind as enter into consideration of language, which will be discussed in due course.

The above discussion leads one to anticipate that understanding of how the brain works as a whole would follow in a similar manner on the basis of a much more extensive analysis. In this scenario, there would be a large catalogue detailing each of the improvements occurring during evolution, and the coordinated mechanisms that implemented these improvements. These accounts would be given in terms of agents, and mechanisms involving agents.

Selection of objects in the visual field

Selecting an object in the visual field provides a slightly different kind of example in that a parameter is involved, namely the information that has to be provided to select an object, sent to the visual system so as to cause (part of) it to output selected information related to the kind of object concerned. Many different variant systems are probably involved, of which we focus on just one. Here we assume that there is an agent corresponding to the object, which is turned on for dealings with that kind of object and

may have different states corresponding to the way in which the object is of relevance at any given time. The agent for the object is created, by a neural system designed for the purpose, the first occasion on which attention is given to the object, and is trained to do various things by modifying its connections with other agents. One way in which an object may be relevant is the one discussed, i.e. as something to be selected in the visual field for the benefit of other agents, and this can be achieved by a mechanism involving a request to try to select the object, that is to say to send signals to the visual system that cause visual information relating to the object best to be filtered out. A typical system for achieving this would be one that linked the object-agent to the neurons of the visual system that are firing when attention was given to the object, the mechanics involved being such that sending signals to such neurons would enhance their firing rates on a future occasion.

Approaching objects and fetching and moving objects

It is assumed that since approaching and fetching objects are such universal kinds of action that evolution has caused systems specific to this activity to come into existence. This means in particular that there is a process that emits 'request' signals for such activities and leads to processes related to learning how to perform them to come into existence. Approaching has already been discussed; fetching is somewhat more complex as the system must remember where to go after the object has been approached and picked up. There may be a special register implementing the concept 'here' to store this information. Before approaching the object a marker for 'here' is linked to the 'here' register, and after the object has been collected the 'here' agent is activated to find out what it has been linked to and this is used as the target of the next approach. For moving, a different register 'there' can be used to define where the collected object should be taken to. These mechanisms are speculative but are suggested by the fact that they are simple and conform with our intuitions that we do have concepts such as 'here' and 'there'.

Planning and memory

In the above we have a simple form of planning operation, in that linking registers such as 'there' to other agents (an assignment operation) will set up a projected process. Such assignment operations will in general originate in requests generated by other agents, which may for example respond to an 'in the wrong place' exception by calling an agent and setting its parameters appropriately. This process can only be one for immediate execution, since the registers used will have their contents changed regularly. A more advanced planning operation would

create an agent specific to the given task, complete with its own registers. This requires a system equivalent to the thread construct in object-oriented programming, which will be discussed later.

It is worth noting at this point that more than one kind of memory has featured already in this discussion; namely that of a module that may be held on for a time, and a memory mechanism that involves linking of two agents. Threads form a third kind of memory mechanism. Memory features differently in this approach than in the neurosciences which treat memory empirically: it features just as another kind of mechanism required to make a process work properly.

Complex operations

Approaching, etc. are almost certainly innately prescribed (although not specified in detail). Other operations, such as those involved in driving a car, are clearly not. Nevertheless these may be acquired by innately specified systems, which can for example train new agents to perform sequences of actions, possibly specified by language. Here again the point can be discussed in evolutionary terms: hardware that can perform particular kinds of concatenations increases the flexibility of the system increases its flexibility and is so expected to be favoured by natural selection.

Threads of control

Threads are a mechanism used in modern computer languages (see for example Niemeyer and Peck 1997, chapter 6) and which in a form adapted to neural hardware appears to be a necessity for high level cognition. They are a kind of software object that has the ability to ensure continuity over a period of time. Their defining feature is that they can be told to wait and cease operation for a period of time until notified that they should restart. Necessary state information is stored before the thread hibernates. The process is somewhat like a subroutine call, but is applicable even if a number of threads are running in parallel at one time. A primitive process that achieves the effect of a thread is one which involves a memory register such as a destination. If one is distracted for a moment, one simply recovers the information as to what to do from the relevant register. This process cannot work reliably in situations which are either complex, such as language processing, or involve distant planning where such information will soon be overwritten by other processes.

Instead of this mechanism a thread mechanism could be used instead, as long as there were a specific cue to start up the thread again in the anticipated condition. An example of planning using threads can be illustrated by supposing we decide to do a sequence of actions driven by agents A, B, C We create a

new agent of the desired kind and teach it to activate the agents in turn (but in a special mode where the actions of the agents are inhibited). We also teach it to be sensitive to signals being received from the agents which they might emit to inform other agents of their state. This is the planning phase. In the active phase the higher-level agent starts up a agent and goes into a wait state itself, starting up again when the agent transmits a signal saying it has finished. We may think of this as analogous to an executive sending at which point he looks up a list to determine what the next job to be started should be.

The problem of language

In this section it will be shown how the concepts developed here account in a natural way for the observed features of language and for its effectiveness. The account develops in a systematic way ideas due to Josephson and Blair (1982).

We begin by characterising communication in very general, abstract terms, as follows. Communication has two components: for the sender it is the act of sending a message correlated with the requirements of the overall situation, and for the receiver it is a matter of interpreting of the event consisting of a message being sent. This account applies for all levels of sophistication of messages, from the crudest to the most advanced. Evolution of communicative abilities is a matter of finding more and more advanced ways of performing these two processes (especially in regard to the organisation of events in the communication channel), and this can be studied in terms of the features that successively emerge.

In the crudest form there are only fixed sounds conveying specific messages; that is to say a particular sound is emitted in response to particular kinds of situation in a way determined innately, and the receiver learns appropriate responses, a component of these responses possibly being innate.

Going beyond this involves the capacity for one individual to create a new signal for a new situation, which other individuals can learn an appropriate response to. A number of auxiliary mechanisms can enhance the outcome of such a process, in particular for the sender to be able to give an indication that the receiver's response is the one desired, and for the signalling system to be one that makes clear distinctions. The latter favours development of the use of specific types of sound like the human phoneme system, and the means of recognising the phonemes being used. Then again, progression to the use of word-like phoneme strings makes a wider range of standard signals available to the intercommunicating community.

As described thus far, the system relates to 'personal languages' where in principle each person uses a different language of her own. Imitation mechanisms

that tend to lead individuals to copy the words used by others in a given kind of situation lead to the a more useful scenario of a language shared by an intercommunicating group.

The basic format for communication at this stage is the emission of the single word-like entity best fitting the situation, interpreted by the listener so as to give an appropriate response. In the most advanced stage there is an agent corresponding to each word, which is turned on to speak and also while listening to diagnose which words are used. The single-word stage is also found during the development stages of human linguistic behaviour. Training mechanisms involve a variety of mechanisms, an example being linking an agent for the word for an object together with one for dealings with that object.

The concept of requests being made that the individual tries by trial and error to conform to is relevant in accounting for the details of some of these processes. For example, interpretation involves the request 'find out what it is best to do (in general, which system to activate) when a person emits this signal', and selecting a signal involves the request 'find what signal works best in the given context'. More subtly, the latter may involve an 'exception' for new signals or ones that are not understood, which can be caught by processes such as pointing or explaining in words. At the time, explaining by pointing after having failed with language does nothing that pointing by itself would not have achieved, but if the listener has a process that links the two it may be possible for the explanation to be omitted the next time.

Beyond the single-word stage: the technology of constructions

In the light of the above, we now assume that there was a stage in the evolution of the language ability where there was a lexicon, and communication consisted of emitting single words from the lexicon, which elicited an appropriate response. This would be a powerful system but limited in its capacities to the kinds of messages that could be communicated with a single word from the lexicon. The system as described could be stretched by the speaker emitting more than one word. The listener could hold the words in memory and attempt to find an appropriate response.

What is involved in 'finding the appropriate response'? Take the case discussed before where a process is described in terms of a particular agent using particular registers for the parameters needed. The appropriate response involves activating the right agent, and linking the agents corresponding to the parameters to the appropriate registers. The agents concerned are determined by the individual words, so the listener just has to link these agents to the correct registers. It will be advantageous now to have a

system utilising learnable conventions so that it can tell without guessing which registers to use.

In fact, natural languages have the feature that one can consistently derive the correct registers from rules associated with the head word, which is the word corresponding to the active agent (the one not corresponding to parameters stored in a register for use by another agent). Pinker (1994) gives as examples (p. 114):

- with the verb *frighten*, the subject causes fear and the object experiences the fear
- with the verb *fear*, the subject experiences fear and the object is the cause of the fear

Furthermore, with the active construction (using the active form of the verb and thereby distinguishable from the passive form), a standard order is used, subject-verb-object in the case of English, so that the word order is sufficient to distinguish subject and object. Appropriate hardware, taking the head word (corresponding to the active agent) as key context, could learn the connections, rules and link the two correctly. Why should such a system evolve if speakers do not have systems for producing grammars like these? The answer is that individual speakers probably tend to use particular word orders in their speech, so that it will still be of benefit for listeners to have such a system. Then individual conventions can become universal for a group by similar mechanisms as have been discussed for the case of word meanings.

Threads, trees and transformations

Such a system for language is quite primitive since it cannot account for tree-like structures. The fact that embedded constructions can be very long without the listener losing track of where the sentence is going strongly suggests that thread mechanisms are involved. When a word does not fit in, the listener starts a new thread and tacks information on to that thread. This process mirrors what *speakers* do, since embedded constructions originate when for example an object cannot be indicated by a single word and a phrase must be used instead, the generation of which demands starting a new thread. As discussed for the case of planning, a given thread goes into a wait state if certain events occur, and waits till another thread notifies it that it can start up again.

Another kind of problem a general theory has to explain is the existence of the transformations originally conceived as devices to connect together a 'deep structure' conforming to the rules of grammar and the transformed surface structure of the speech output (Pinker 1994, pp. 120–4). An example is the sentence 'What did you put in that box?' where there is a gap after 'put' that would normally be filled by a noun phrase, as in 'Did you put my screwdriver in

that box'? Here we have a sentence that it is perhaps natural to make, following a convention of putting a question-word first and then indicating what the question is about, but it is more difficult for the speaker to disentangle because he has to work out that the thing that the speaker wishes him to identify is the thing that was put; a structure must be instantly built to reflect that fact, in just the same way as if instead a statement had been made such as 'I put your screwdriver in that box'.

The comprehension of such statements can be tentatively understood in terms of the thread mechanism. The parsing apparatus might initially consider 'you' in the sentence 'What did you put in that box?' to be the first word of a noun phrase that would be the object of 'did' (as in the question 'What did that?'), but this hypothesis would have to be abandoned when the word 'put' was encountered. Instead, a thread started with 'what did?' would have to be put into a wait state until some event occurred to notify it that it should start up again. The event that hypothetically does this is the absence of a noun phrase after 'put' (which in English takes an obligatory object), so then the information in the 'what did' can be added in at that point. The 'did' merely confirms the past tense while the 'what' supplies the noun phrase required by the grammar, and its existence is an arbitrary fact contingent upon the origins of that particular structure in the language, since a priori the alternative 'What you put in that box?' would convey the speaker's intent equally well, and may indeed be used by some speakers.

Origin of syntactic categories

It appears, then, that consistent manipulation of threads can in principle account for our ability to manage complex input and construct structures which reflect the semantics (equivalently, agent-complexes able to do the tasks entailed by the meanings in the utterances concerned). However, the above discussion ignored the question of how we know what fits in in any given situation, which is crucial in determining how the components of the threads are to be fitted together.

The situation resembles that of a person accumulating a stream of objects into boxes. The contact with visual metaphor may be increased by imagining that these objects have their own characteristic shapes, and that only certain sequences of shapes are admissible for any particular box. Furthermore, the contents of a completed box have a characteristic shape also which determines into which boxes these contents may be put. This account is the equivalent of grammatical constraints governing the structures that can be built, such as (Pinker 1994, p. 197):

"A noun phrase can consist of an optional determiner, a noun, and an optional prepositional phrase."

The receiver of the stream then simply puts its elements into the uncompleted boxes and takes a new empty box if something comes that will not fit into the current box. When a box has been completed it is treated as if it were input for the box whose filling process was interrupted. In accord with the discussion of transformations between deep structure and surface structure given above, some items are kept separately until a space opens up to contain them. (The problem of anaphora has not been discussed, and may be related to threads being kept active for the purpose of making anaphoric connections for a time afterwards, or to the use of special registers such as those hypothesised for concepts such as 'here' and 'there').

In the above account, some kind of input buffer mechanism is needed to hold the incoming information until it is decided by the various agents looking for valid structures which 'boxes' the individual items are to go in to; i.e. which threads the items are to be linked to.

Just as specific circuitry facilitating imitation of one person's use of words can transform word conventions from conventions specific to an individual to conventions specific to a community, other kinds of imitative mechanisms can facilitate the transformation of a single individual's idiosyncratic groupings of words into the grammar of a linguistic community. Newly devised constructions will tend to propagate into a community in proportion to the extent that they are simultaneously useful, and easy to interpret against the background of possible incorrect interpretations.

The question arises what training schemes are used by individuals to discover the classifications used by speakers of a language in general. There are two main ways by which the types might be diagnosed by a system capable of representing types, namely through semantic cues and through syntactic regularities. In the simplest constructions the syntactic types reflect the semantics, but this close relationship is violated by more complicated constructions such as the gerund, which syntactically is a noun while being semantically an action. However, networks of the kind discussed by Elman (Churchland 1995, pp. 137–43), which learn acceptable word orders and develop hidden layer structures related to grammar, might be able to extend the categories initially learned on semantic grounds to more general ones. Elman et al. avoided including modules with explicit type representations, but a more realistic model is likely to include specific structures dedicated to this activity. The learning networks then become devices that output the types expected in a given context, and this provides a natural mechanism for the system to postulate tentative *types* for previously unknown words or phrase structures.

What comes after understanding?

The above account is presented as an account of what happens when we listen and understand. A structure of linked threads is built up as a result of solving the problem of which structures have been employed by the speaker, and interpretation rules, which can be applied by agents corresponding to them, are available, as discussed in the section entitled 'beyond the single-word stage', to make an interpretation of this structure. This interpretation indicates the role of every item that has been linked to the individual threads. The overall form of the information in the sentence has been discovered but the specifics of how to use the information, which fall outside the domain of language itself, have to be learnt by trial and error, building up through the solution of simple problems a collection of agents that can be utilised again in more complex situations. The details of this lie outside the scope of this paper.

The nature of language

The picture we ultimately arrive at, starting from what is known about the regularities of language, is one involving a small number of specific mechanisms that help develop the process of communication. This picture is neither that of Elman et al. (1996), who claim that no mechanisms exist that are specifically appropriate for language, nor one that takes Universal Grammar as absolutely fundamental. There are simply powerful mechanisms that drive the process of language along, and what we hear is what they give, which may for a variety of reasons tend to conform to certain regularities such as those encompassed by Universal Grammar. Every now and then, a speaker, frustrated by the inability of the existing language to convey the meaning he intends, throws an exception and tries to catch it by inventing a new word or by using a hitherto illegitimate construction. He may then have to catch the speaker's inability to make sense of his invention by means of various devices of explanation. Some of the inventions are useful enough that they propagate among users and become a part of the language of the community.

Conclusion

This paper has attempted to show how a synthesis of a number of established ideas can provide powerful insights into how the mind may work, including advanced cognitive processes such as those of human language. A considerable degree of fit to the model and observation is obtained in a fairly straightforward way, which I hope will encourage further exploration.

Acknowledgements

I am grateful to Profs. N.A. Baas and A. Ehresmann for discussions concerning their own abstract approaches to the problem of the complexity of the mind, and to Dr. H.M. Hauser and Dr. D.G. Blair for earlier discussions of the mechanics of development, out of which the present approach originated.

References

- Back, T. (1996) *Evolutionary Algorithms in Theory and Practice*; Oxford University Press
- Baas, N.A. (1994); *Emergence, Hierarchies and Hyperstructures*; *Artificial Life III* (ed. C.G. Langton, Addison-Wesley (pp. 515–537)
- Brooks, R.A. (1986); *A Robust Layered Control System For a Mobile Robot*; *IEEE Journal of Robotics and Automation*, Vol. 2, No. 1 (pp. 14–23)
- Churchland, P.M. (1995); *The Engine of Reason, the Seat of the Soul*; MIT Press
- Ehresmann, A.C. and Vanbremeersch, J.–P. (1987); *Hierarchical Evolutive Systems: a Mathematical Model for Complex Systems*; *Bulletin of Mathematical Biology*; Vol. 49, No. 1 (pp. 13–50)
- Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A.; Paresi, D. and Plunkett, K. (1996); *Rethinking innateness: A Connectionist Perspective on Development*, MIT
- Josephson, B.D. and Blair, D.G. (1982); *A Holistic Approach to Language*; e-print at <http://www.tcm.phy.cam.ac.uk/~bdj10/language/lang1.html>
- Josephson, B.D. and Hauser, H.M. (1981); *Multistage Acquisition of Intelligent Behaviour*; *Kybernetes*; Vol. 10 (pp. 11–15)
- Minsky, M. (1987); *The Society of Mind*; Heinemann
- Niemeyer, P. and Peck, J. (1997); *Exploring Java*; O'Reilly
- Pinker, S. (1994); *The Language Instinct: the New Science of Language*; Penguin
- Pinker, S. (1997); *How the Mind Works*; W.W. Norton