

# CONSTRUCTIONAL TOOLS AS THE ORIGIN OF COGNITIVE CAPACITIES

**Brian D. Josephson**

Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, U.K.

## Abstract

It is argued that cognitive capacities can be understood as the outcome of the collective action of a set of agents created by tools that explore possible behaviours and train the agents to behave in such appropriate ways as may be discovered. The coherence of the whole system is assured by a combination of vetting the performance of new agents and dealing appropriately with any faults that the whole system may develop. This picture is shown to account for a range of cognitive capacities, including language.

## Keywords

Development, cognition, agents, neural networks, tools, paradigms, domain specificity, issue resolution, language.

## Tools and development

The aim of this paper is to explain how complex mental capacities might arise naturally from a physical system like the brain. It is postulated initially that the nervous system incorporates a collection of *tools* that serve to create a collection of agents which cooperate together to create the capacities that we observe. Tools are best envisaged as systems that under appropriate circumstances systematically utilise existing capacities to create new ones. For example, while in a standing position, the existing capacity to stand in balance is utilised by one of these tools to attempt, through some systematic process which we shall not consider in detail, to take a step. When the process is successful, or nearly successful, an agent system is created to learn the process, so that future activation of the agent will repeat what has been learnt.

Initially, by virtue of the specific links between the agent and the rest of the nervous system, the learnt

process is activated only in circumstances similar to those for which the skill was learnt. It is hypothesised however that, as a result of some mechanism capable of sustaining the activation of an agent, the capabilities of the agent are then investigated in other circumstances. The process may then fail totally, or be able to be corrected by some other mechanism. Through appropriate adjustments, an outcome is then arrived at whereby the agent becomes active when called upon to act only in a range of contexts where the performance has been found to be successful in the past, and is inhibited otherwise, implying that in general agents will behave in the desired manner.

In physical terms, what the tools are doing is constructing mechanisms. In the above example, the mechanism for taking a step makes use of mechanisms for moving a leg and for carrying out such adjustments as may be necessary to maintain balance. The tool must be able to access such mechanisms in order to perform its trial actions, and it must be able to connect the agent to such mechanisms in the appropriate manner to endow the agent with the ability to perform such actions when called upon to do so. This implies specific circuitry and corresponding processes for making appropriate connections; the logic of the operation of the tool must be paralleled by neural circuitry whose details relate to the logical details in an appropriate manner. Relationships of this kind between neural architecture and corresponding function are well known and the details are not our concern here; instead, our task is the question of how a suitable collection of tools may be able to construct the mechanisms we observe, in the case of advanced capacities such as language.

In the case of walking, we discussed the case of a tool that performed a process leading to the capacity to take a step and retain balance. Our thesis is that all the capacities we have can be built up cumulatively, step by step, by means of appropriate tools, in just such a manner. Suppose for instance that we have already acquired a collection of agents, each capable of taking steps in

somewhat different ways. At a higher level of development, specialised tools may investigate which of these agents may be able to achieve more specific goals such as being able to step to a particular place, or approach some object. The various tools thus build up repertoires of activities of particular kinds, each implemented by particular collections of agents that can be activated by other tools to create other kinds of agents again.

It is an important feature of tools that particular combinations of them can produce results of high utility, as already been illustrated with the case of walking, the development of which according to the present proposals involves the application of a range of tools (including some not discussed such as ones that involve generation of the capacity to achieve a vertical position or to stand in balance, or abilities such as taking a single step, and the creation of agents which will take a series of steps rather than a single one, and finally agents which record the activities involved in tracing particular routes, leading to an implicit representation of routes).

## Issue resolution and paradigms

A tool in its most general form can be viewed as a device that may be helpful in resolving an *issue*. Thus tools may be of a variety of kinds. Some, such as those concerned with learning to balance are probably innate, though possibly modifiable by experience. Others may develop through personal learning or by learning from others. An important type of tool has as its basis specification of how to resolve an issue in terms of language, which among other things makes available cultural knowledge and allows knowledge to be accumulated in a culture, so that progress can go far beyond what is possible for a single individual in the span of a lifetime.

An integrated view of what has been said above can be expressed in terms of the concept of paradigm. A paradigm is defined as the collection or pattern of activities of a particular system that can occur in some particular context, wide or narrow as may be appropriate; the paradigm is a function of the system and the context in which the system operates. Radically new paradigms *emerge*, as and when a suitable combination of tools, capable of dealing with the various issues that are important within the paradigm, emerges. in the course of evolution. During development, as the tools encounter the various issues and deal effectively with them, new behaviours associated with new agents come into existence and the paradigm is thereby extended. Such extensions, while

increasing the scope of possible activity, may be associated with problems, but further development on the basis of the tools that deal with the relevant difficulties can lead to the resolution of the problems concerned, or alternatively to the inhibition of the agents responsible for the activity associated with the problems, or again to the avoidance of the situations where the unavoidable problems occur. In this way it is ensured that the system always behaves in an integrated manner.

A question to which a definitive answer can be achieved only by investigation of the details is that of which agent collections will suffice to deal with most situations encountered in practice. This question is bound up with that of which explorations are useful and will lead to a sufficiently representative sample of situations being investigated. In some situations, random exploration will be sufficient, while in others more focussed exploration may be necessary. This may involve guidance by more experienced members of a culture.

In any event, the picture hypothesised here is that the right kind of tool (leading to an appropriate kind of exploration, and responding to the right kind of cue, leading in many cases to particular issues being resolved with the creation of corresponding agents) does lead to the expansion of paradigms (patterns of behaviour) in a way corresponding to what is observed, and that these tools operate cumulatively to produce the totality of observed behaviour.

It should be noted that these tools do not need to specify in advance the precise behaviour that should arise from their actions in a task; specification of the explorations and the means to resolve issues being all that is needed in many cases, with the population of agents that result containing the information of which particular way the issues were resolved during these explorations. This is well illustrated by a case such as rock-climbing, where it is clear that most of what is learnt can be regarded as the resolution of certain general issues such as where and how to step, how to stand in balance, and how not to fall. Even ice-skating involves resolving the same issues, under the considerably different circumstances that obtain for this process!

## Higher order skills

### (i) *planning*

It may seem reasonable to hypothesise that some collection of specific tools can lead to the cumulative development of skills concerned

directly with what is perceptible to the senses, but less clear how higher cognitive capacities might be encompassed within such a framework. To see how this might be done, let us postulate that besides the systems concerned directly with action and perception there exist support systems containing agents concerned with representations that do not have to refer to the immediate situation. These systems have to derive their activity in the first instance from something and we assume (apart from possible innate routines) that they derive initially from action and perception but they can subsequently become temporarily divorced from action and perception. (Temporarily is the key word here; if they were to remain permanently divorced from action and perception they would have no value.)

At the very least, we need to have tools that can create agents that correspond to significant elements of the sensory world, and can link these back again so as to reflect back on elements in the sensory world, e.g. to repeat a represented action or to identify a represented object. They must also have a capacity to link together these systems representing aspects of the sensory world in ways that can be used imply important relationships in the sensory world. Thus there can come the ability to abstract out and import into the supportive system phenomena in the sensory world, and then later export them again so that events such as pulling on something to make a noise can be repeated. Such an elementary process provides a very simple paradigm of support of activity by the representing system, and one can then go on binding together agents in the supportive system in ways that have more complex implications in the sensory world. For this to be of relevance there needs to be some process which can switch between a connected mode and a disconnected mode, so that constructs created in the disconnected mode can be tested in the connected mode.

This is just a sketch of the possibilities, but in summary we can say that certain aspects of the world are representable in a symbolic form, in terms of which there can be processes for the creation of useful structures. Random symbolic activity is liable not to lead anywhere, but by trial and error one learns about what useful possibilities there may be of this kind, and builds up corresponding collections of agents capable of generating representations which can be used in activity. As noted earlier, guidance by others may be relevant in support of acquiring such paradigms.

#### *(ii) language*

Such a process is more explicit, and can be followed through in more detail, for the case of

language, since we know more about the various patterns and processes involved in language. It is found that the range of languages has considerable uniformity, described by so-called universal grammar. It can be argued (Pinker, 1995) that this uniformity makes language easier to acquire, and is also indicative of the particular processes whereby language functions effectively. It will now be shown how the concepts developed here can give an account that is parallel to Pinker's but is more precise, and indicative of mechanistic details.

In the first place, what language is can be related to what language does, which can be defined as exchanging information in such a way as to facilitate outcomes desired by the speaker. From the listener's point of view, this is a special case of interpreting incoming information, special in that the information originates from a sender who has certain views on what the receiver does with the information imparted, so that the information can be considered as directed towards some intended conclusion, which circumstance may not be the case with information in general.

Language can be viewed primarily as a way of systematically encoding certain neural structures, in such a way that a decoding process creates corresponding structures in the mind of the receiver. Speakers gradually learn which structures are relevant to encode in order to fulfil particular intentions, a simple case being that of naming a desired action. If a speaker has a process which creates a name for a desired action on the listener's part, and the listener can initially associate the name with his representation of the action, then on subsequent occasions the speaker can use the name again and get the corresponding action without any further training. What we have here as an encode-decode pair which complement each other in that the action of the two together is to produce a suitable correlate to the initial representation of the action. Note that the decoding process does not necessarily have to be derived by investigating all theoretically possible referents of the name; if there is some alternative process which points to the intended referent (e.g. the presence of some correlation, or some act of pointing or other means of indication) then the association can just be learnt; in other words there exists a specific tool facilitating the learning process.

It is useful to consider the speaker and others as using a paradigm, viz. a pattern whereby name and referent are correlated (in a way that is in general context dependent). The listener, by observing the paradigm (possibly taking into account the context dependence) can create a decoding agent by following some recipe (i.e. given the name, create

the observed reference). In addition, the listener can, using a slightly different process that works in the reverse way on the information available, create an *encoding* agent which, given the representation of the reference, generates a process for the name. We already see the utility of hypothesising tools that create agents according to some specified process.

The kind of process just discussed (i.e. having tools that create agents for coding or decoding systematically on the basis of the paradigmatic regularities present in the linguistic environment) can be seen to apply at other levels also, and can be seen to be the rationale for universal grammar. These paradigms reflect the ways in which the text conveys information, e.g. using the so called X-bar structures (Pinker, 1995, 107) as ways of indicating which constituents occupy which roles; or using phrase structures with elements of particular types in a particular sequence to indicate where the phrase boundaries are in order that the appropriate hierarchical structures can be constructed. The decoding apparatus needs to have an assembly of agents that operate in turn at appropriate times.

The question as to how all this coordinated machinery comes into existence is a complex one. The basic idea seems to be that tools develop the ability to carry out systematically things that are beneficial and occasionally happen by chance. It has already been indicated how a tool designed specifically to look for correlations between speech and the speaker's intentions could facilitate the learning of uses of names. Again, naming itself (making a sound correlated with one's intentions) may happen by chance as a result of existing neural connections but a suitable tool mechanism could make it happen more consistently. We may assume that the conventions of universal grammar arise in a similar way: certain consistent patterns of speech may occur naturally and also be easy to decode, in which case it will be beneficial to have a tool specific to creating speech patterns according with these general patterns, as well as a tool to decode such patterns.

With simple languages based on simple grammars (such as were used in artificial intelligence simulations such as that of Winograd (1972)), it may possible to see explicitly how processes for creating agents reflecting the paradigm would work. Real languages appear to be much more complex; instead of explicit rules that state definitely that a particular process is to apply we find fuzzy rules, as well as the resolution of ambiguities in terms of context. This is not hard to understand on the basis of our picture. At any given time there is a given language paradigm

generated and interpreted by particular constellations of agents (which may vary from individual to individual, even though commonalities are dominant, since different individuals have different environments). Situations may arise that fall outside the current main paradigm and require the tools to be applied to extend the paradigm, for example by inventing a new word or a new X-bar pattern. This implies new agents that may under favourable circumstances, such as the utility of the corresponding paradigms, propagate through the community. New agents are tested to see how well they fulfil their purpose, for example how well they convey meaning. Potentially ambiguous structures can arise (e.g. using a word that can have different meanings in different contexts) that are resolvable in practice if the excitability of agents is fuzzy in nature and context dependent. In computer jargon, much overloading is possible, to a degree related to the possibility of resolution by contextual information.

There is a vast amount of structure and regularity in language, and no attempt will be made here to specify this exhaustively. In general, the details fit with the concepts that have been discussed.

One particular issue that deserves discussion is how much language-specificity is involved in the system or is necessary. The conclusion we are led to is that, as in the discussions of Pinker, conformance to universal grammar simplifies many aspects of language processing. If there are tools specifically adapted to the paradigms of language then everything will work more efficiently. The tools do the kinds of things that neural networks do in general but they do it selectively, and circuits instantiating the required activities of these tools would imply a better functioning system.

Another specific aspect of universal grammar is one worth addressing explicitly, namely the division of grammatical objects into types such as nouns and verbs, which types are used to differentiate between appropriate and inappropriate groupings and thereby help to resolve ambiguities in the conversion of linear to hierarchical structure. The difficulty in interpreting this observation is that grammatical types do not always correspond to semantic types in the expected manner: for example, a gerund has the semantics of an action but its grammatical type is a noun. Now if we think of a word such as 'singing' we can see that in fact there is ambiguity in whether it should be thought of as a thing or an action, since it is both something that people do and something that people hear; the different aspects are fused into one unless we especially wish to discriminate between them. But

when we say something about singing we tend to do such discriminating; for example if the information to be conveyed is that a particular person is doing the singing then the emphasis is put on the action aspect, while if we want to say that the singing is loud then the emphasis is on the sound, which is treated as a thing since one hears it.

The crucial point is that one seems to be forced to make such a distinction, as assists the determination of structure, but the origin of this distinction is probably related to the different ways actions and objects are represented in the brain generally. Here the relevant tool (for detecting groups) is one which takes note of which areas of the brain are active, and which in creating an agent from a group tries to respect existing patterns.

The general pattern in the above has been the same as other instances that have been discussed: specific tools lead to the paradigms of activity being gradually extended. Certain characteristics of the resulting agents make this activity tend to be useful; thus the tools have a certain potential that can be fruitfully realised. As agents accumulate, the activity that they cooperate in becomes more and more complex, but the vetting of new additions to the system and of the overall activity of the system ensures that it remains useful and in control (ideally, of course; we know that in human societies, such regulatory activity does not always work very well).

## Discussion

We have developed the idea that a collection of special-purpose tools, adapted to cope with a range of issues arising in various situations, may play a very important role in the establishment of cognitive functioning. The effect of the operation of a given tool is to create an agent capable of performing a specific kind of task in the given context. Provided that agents are selectively located in regions of the nervous system in such a way that location is indicative of their functions, other tools can locate them as required so that a system of tools can develop agents in a cumulative manner in accord with appropriate rules or algorithms. It was argued that processes such as walking, and equally the use of language, can be implemented by collections of agents of various types, the agents being constructible through a process of attempting comparatively simple tasks by trial and error. These agents resolve particular issues in simple situations, and it is proposed that skills can be acquired in more and more complex situations merely by resolving a certain set of issues in those situations where such issues can be

resolved. As more and more agents are added, the domain of proficiency within a given paradigm progressively expands.

Interacting-agent models of various kinds have been proposed by various authors, for example the society of mind model of Minsky (1987) and the neural constructivism approach of Quartz and Sejnowski. (1997) and of Elman et al. (1997). Neither of these approaches give clear indications of how development might be organised. The idea of a synergetic interaction between tools, agents and the resulting paradigms, the consequence of considerable specificity in the design of the nervous system, has the potential to resolve such issues. Granted, the ideas that have been proposed here are speculative and intuitive in nature, but support from them may be obtainable by means such as detailed experimental investigation combined with analysis of the phenomena concerned. Indeed, the analyses of Karmiloff-Smith (1992) already provide strong evidence for the existence of mechanisms of the kind proposed here in particular realms of activity. It is tempting to suppose that similar concepts apply to further aspects of cognitive functioning, including intellectual activity, with for example tools that operate with constructs such as propositions, inference and truth, fruitful paradigms in these areas being developed by such tools over the course of time.

## Acknowledgements

I am indebted to Nils Baas, Andrée Ehresmann, Burghard Rieger and David G. Blair for discussions related to the ideas proposed in the above.

## References

- Elman, J.L. et al. (1997); Rethinking Innateness: A Connectionist Perspective on Development, MIT.
- Karmiloff-Smith, A. (1992); Beyond Modularity: a Developmental Perspective on Cognitive Science, MIT.
- Minsky, M. (1987); The Society of Mind; Heinemann.
- Pinker, S. (1994); The Language Instinct: the New Science of Language; Penguin.
- Quartz, S.R. and Sejnowski, T.J. (1997); The neural basis of cognitive development: A constructivist manifesto; Behavioral and Brain Sciences, Vol. 20 (4): pp. 537+.
- Winograd, T. (1972); Understanding Natural Language; Edinburgh.