Symbols, Signs and Cognition

David Davenport

Bilkent University
Department of Computer Eng. & Information Science
Ankara 06533 - TURKEY

Technical Report BU-CEIS-94-18

Symbols, Signs and Cognition

David Davenport
Computer Eng. & Information Sciences Dept.,
Bilkent University,
Ankara 06533 - TURKEY

email: david@bilkent.edu.tr
david@trbilun.bitnet

Abstract: Cognition is inextricably linked with the concepts of signs and symbols. The classical, computational, view of cognition, explicitly states this in its "physical symbol system hypothesis." Its major competitor, connnectionism, while claiming to be subsymbolic, implicitly acknowledges both signs and symbols. Unfortunately, both theories have their problems. The computational paradigm fails to explain where symbols come from and how they acquire meaning. The connectionist paradigm fails to provide sufficient representational power and anyway is based on the discredited associationalist philosophy. While the problems are different, the precise nature of the two approaches and the essential difference between them, remains something of an enigma. This paper outlines a theory of "symbols" which not only appears to provide a clear basis for distinguishing between the two paradigms, but also points the way to a new architecture which may help solve the mystery of cognition.

Keywords: Artificial intelligence, symbol systems, connectionism, cognition

Introduction

According to prevailing theories of cognition, the relationship between people, signs and symbols is very intimate. Not only do individuals communicate using signs and symbols, signs in the form of bodily gestures such as pointing, and symbols in the form of words and pictures, but they actually think using them too. Such theories suggest that cognition itself literally *is* the storage and manipulation of symbols and/or signs.

The most prominent cognitive theory, the classical computational paradigm, takes as its basis the physical symbol system (PSS) hypothesis which says that, "A physical symbol system has the necessary and sufficient means for general intelligence" [1]. In other words, intelligent agents (such as people) will, upon examination, turn out to be physical instantiations of formal symbol systems. That is, they will store and manipulate symbols in essentially the same fashion as does a modern digital computer.

The classical theory's main rival, connectionism, claims to be "subsymbolic" and therefore does not explicitly acknowledge symbols, but symbols and signs are quite clearly implicit in its make-up. Unfortunately, both the connectionist and symbolic paradigms present problems when viewed as architectures of cognition. The former, according to Fodor & Pylyshyn [2] lacks representational structure and is based on the discredited associationalist philosophy. The latter, given Searle's Chinese Room Argument [3], simply manipulates symbols without any idea of what they mean, where they come from or how they are represented. Although each seemingly presents a different set of problems, the difference between the paradigms is actually far from clear. Afterall, both are ultimately implemented as electrical signals on bits of wire. Furthermore, they have been shown to be functionally equivalent, neural nets being (at least theoretically) Turing-compatible and anyway commonly simulated on symbolic machines.

This paper sets out to explore the nature of signs and symbols, and their relation to cognition. Developing a sound conception of signs and symbols will not only provide for a better understanding of the status and failings of these two competing theories, but will also show a way to overcome their limitations. The following sections look at signs and symbols from the point of view of communication, storage and recall/matching. The symbolic and connectionist paradigms are then re-examined from this new perspective, and an alternative theory of cognition is proposed.

The Nature of Symbols

A symbol is a non-natural sign or mark which can represent or stand in place of some thing, some concept. Common symbols are letters, numerals, words, flags, etc. A sign, in turn, is something which stands for, signifies, indicates or signals something else. Signs may be natural (e.g. smoke) or artificial (e.g. an exit sign), and may also vary as to the degree of reliability with which they indicate. Natural signs are rarely guaranteed indicators, relying heavily on our ability to interpret them, for example, black clouds are a natural sign which sometimes indicates rain, while smoke generally indicates fire, a thumb-print, presumably always indicates a particular person and the sun setting always indicates the approach of night. Artificial signs, however, can be as reliable an indicator as desired, an exit sign always indicating a way out, the word 'cup' indicating a drinking vessel, the plus sign addition, and so on. Notice that the term symbol may be used interchangeably in these latter cases.

Abstract signs/symbols may be physically instantiated in any number of ways. Indeed, anything the presence of which, at a particular time/place "conveys" some kind of meaning or idea, and the absence of which fails to induce the very same meaning/idea, can be taken to be a sign or symbol. The meaning/idea conveyed may be extremely simple or very complex, but it is not inherent to the sign. In order to be appreciated and accepted as a sign/symbol the meaning must already exist in the recipients mind and some sort of connection must relate the sign/symbol to it.

Communication

To clarify these ideas, consider how signs/symbols can be used to communicate meaning (this from a theme originally developed in [4]). A "signing" mechanism is implemented by a man standing on the top of a hill with a flag in his hand. Assume for the moment that he is allowed to hold the flag in only two positions, down beside his leg and up above his head. Now, when we observe this man from afar, what information (knowledge/meaning) can he convey to us? Unfortunately, none! Only if we adopt some "convention" can we receive any meaning from our "signer"; if, for example, the man raises the flag above his head only when he detects the approach of an enemy. Notice, that we could adopt any number of similar conventions, e.g. raise flag only when you observe the colour yellow in the valley beyond the hill, or only when you feel the wind on your face, or only when you see the man on a hill beyond yours raise his flag, etc. Alternatively, we could adopt the inverse convention, such that, for instance, the flag is only lowered if an enemy is seen approaching.

Independent of whichever convention is chosen, if the man is not signalling the approach of an enemy, we would naturally tend to assume that an enemy was not approaching. However, this is not necessarily the case, the man may simply be unaware of any impending danger (or he may have been taken by surprise, tied up and hence be unable to signal). In fact, the world simply does not naturally "emit" negative information which could be detected and used to signal such things (only language provides this facility). To infer "not" something, it is necessary to rely on positive information provided by sources known to occur only in the absence of the original signal. For the moment we will leave open the question of how it is possible to know/learn this.

Assume now that we wish to be able to convey, not just a single message, but any number of messages. One way to achieve this would be to add more men with flags. Since each man can convey only a single message, we would require as many men as we have messages. Of course, keeping all these men fed and clothed would be expensive and we may wonder if there were not cheaper alternatives. There are several possibilities,

> (a) Employ a single man and get him to signal each possible message in sequence. Doing this reliably requires great care to synchronise the transmission and reception of messages, indeed extra messages and protocols will be necessary to ensure that no errors are made.
>
> (b) There may be a lot of redundancy in the system, such that certain sets of messages always occur together, while others never occur. Based on this knowledge we could replace the group of men corresponding to a set of

concurrent messages, with a single man, instructing him to raise his flag only in such case that all the sub-messages he encodes are actually true. Depending on the redundancy present, this could reduce considerably the number of men required, while still allowing all messages to be transferred concurrently.

(c) Alternatively, the redundancy may be such that certain sets of messages can be shown to be mutually exclusive. In this case, only one of a particular group of messages will need to be conveyed at any instant. This knowledge allows the message group to be encoded so as to reduce the number of men needed. Using a binary coding scheme, for example, would decrease the number required to only log base 2 of the number of possible messages.

(d) Allow the man to hold the flag in many different positions, each one signalling a different meaning. Our ability to detect differing positions will obviously limit the number of messages which can be conveyed successfully by a single person using this technique - although the entire message set could be divided between several men as before. Plainly, using this method means that, at any instant, only a single message can be conveyed, thus restricting its applicability to situations where messages can be shown to be mutually exclusive as in (c).

Techniques (c) and (d) then, rely on the original messages being mutual-exclusive, however, determining such redundancy in a natural environment is a difficult, if not impossible, task. Technique (b) is similarly troubled, in that it requires us to show that certain groups of inputs always occur together. But how can we know or determine that particular groups of naturally occurring signals will always be in conjunction or disjunction. Unless we observe the environment for all time, an impossible and pointless exercise, the answer is obviously that we cannot know, at least not with any certainty! That we can and do use techniques (c) and (d) is due to the fact that, in certain special cases, we can "fix" the environment such that we know that signals must be mutually exclusive.

Finally, observe that however messages are communicated, if more than one man is involved, then it is imperative that we know which man is transmitting which message and when. In other words, the spatio-temporal sequence of messages is critical to making sense of the communication.

Storage

Consider now, the situation whereby, rather than simply communicate messages, the requirement is to actually store them. Retaining the set of inputs so that they are available to future 'computations', is surely a minimum requirement for any sort of intelligent cognitive behaviour (if an agent cannot remember the past it is doomed to repeat it, for better or worse!) In remembering the input set, it is important to know which particular input is related to (caused) a particular stored memory. If this information is not available, the same pattern of signals occurring on a completely different set of inputs would be recognised as being identical which is obviously nonsense.

The problem then, is how to store the entire set of incoming signals, on a more or less continuous basis. Initially, let us view time as discrete and record inputs instant by instant. In the simplest case we must "remember" (record/store) the 'state' of a single "signaller" (the man on the hill with a flag, who now acts as the input to our system). There is only one solution, we must allocate another man whose function is to observe the input (man on the hill) and, at a designated moment, "record" the input's state by placing his flag in an identical position and keeping it there. Obviously, we will require a new man for each time point. Note also that we really need some way of "knowing" which man corresponds to which instant of time, for without this information we have no way of reconstructing the input sequence (hence many different input sequences will appear to be identical, which is patently undesirable!)

Extending this example to the case of multiple inputs is relatively straightforward. For each time instant, simply allocate one new man for each (input) signaller, with orders to record the state of only their particular (allocated) signaller. To ensure that no confusion occurs they should stand one behind the other with their free hand (the one not holding the flag) on the shoulder of the man in front of them. We have, in effect, 'copied' the set of inputs, the required causal information being implicitly retained in the men's positions within the line. In this way then, we have a line of men for each time point, and again require some means of determining (fixing) their temporal sequence. One way to achieve this may be to allocate yet more men, this time without flags, whose job is to "link" pairs of lines (successive time points). They could do this by holding the free hand of the man at the front of one line, and placing their other hand on the shoulder of the man at the rear of the next line in the sequence. In this way both temporal and spatial information could be retained.

An alternative method of recording the input set would be to allocate just one man per time point, with instructions to draw arrows on the ground in front of him, pointing to only those signallers whose flags were up at the designated moment (or down depending on the convention in force). Compared with the previous method, this one is very economical in terms of the number of men needed to store the input set. Further savings may actually be possible in the previous method if the environment can be shown to be redundant and hence encodable either conjunctively or disjunctively (see previous section on Communication). The effect of this would be to reduce the number of input signals and thus, correspondingly, the number of men required to record their states. Such schemes would not affect the second method since it already employs the minimum one man anyway. In this case, however, it may be advantageous to reduce the number of 'links' which each man has to form/record. This may be achieved by allowing each man to link to inputs either directly or indirectly (via other men) so forming a sort of loose hierarchical structure. In essence the set of inputs is being split into several subsets, with one man remembering each subset. These men then act as "inputs" for another set of men, and so on.

For future reference let us term the first method of storage the 'copy' method and the second one the 'link' method. Note that a combination of the two methods is possible. In fact the scheme previously proposed for handling temporal sequence information may be viewed in this light, the extra men (those without flags) effectively act as link-storage. Finally, observe that the pure link-storage scheme itself requires some means of retaining temporal information, and while a similar scheme could be adopted (to 'link' the link-men) this may not be the most appropriate solution. We will return to this later.

Recognition/Matching

Just being able to remember everything is of limited value, what is important is to be able to use this stored knowledge to direct future actions. This requires that an agent be able to somehow match the current situation (current set of input signals) with its memory to locate identical/similar occurrences from the past. Having found matching, or partially matching, situations, the agent can then examine the spatio-temporal information in their vicinity to determine what else to expect. These "expectations" will then help it select subsequent actions - in an intelligent manner? (Note: this assumes a particular view, or philosophy of the world, see [5])

In the link-storage scheme, matching simply requires that each man examine the set of inputs to which he is linked (has drawn arrows to) and, if they all have their flags up, then he too raises his flag to indicate a match. Partial matches can be detected in a similar way, the man raising his flag if one or more of 'his' inputs are valid. Of course this may result in many "matches" being found, some better than others. We might modify our scheme slightly by requiring that each man raise his flag to a level proportionate to the number of his inputs which are actually valid. We could then examine these levels and select the highest one(s) as our "best match".

Finding matches in the copy-storage scheme is, in contrast to the link-storage scheme, quite complex. For one thing there are no conveniently 'spare' flags to use as indicators. Even providing an extra man and flag to each line (group) of men does not immediately solve the problem. Some relatively sophisticated processing is required to compare the current input set with each line of men, and to set the "match flag" only if, for all the men in the line, both the particular input flag is raised and the corresponding man in the line also has his flag raised. As before we could arrange for partial match detection and record the degree of match in the level of the match-flag, finally selecting the highest flag to obtain the best match.

Internal versus External Symbols

Before we move on to re-examine cognitive theories from this new perspective, there is one final and very important point which needs clarification. It is essential that we draw a clear, unequivocal distinction between those symbols which are purely internal to an agent, and those which have a separate external existence. Failure to properly differentiate symbols in this manner, has resulted in much confusion within the literature; for example, the idea that storing the ASCII symbols 'C', 'A' and 'T' in a machine can somehow enable it to gain an "understanding" of the concept cat.

The contrast between the two forms of symbol is clearest in the link-storage scheme. It should be obvious that this method of storing and recognising signals (as described so far) acts much like a simple filter, that is, multitudinous inputs are filtered (or compressed down) into a few "interesting" signals. In this way then, an agent which was viewing say, the letters 'C', 'A' and 'T' written on a piece of paper, would ultimately reduce the many thousands of sensor signals which become active in the process of observing these 'external' symbols, into one or more corresponding 'internal' symbols. Notice that these internal symbols/signals do not have anything of the shape or form of the original letters. There is no question of them having been encoded into any ASCII-like form, rather, they consist simply of a man with his flag raised to indicate a match with a previously encountered input set. It should also be equally obvious that there is absolutely no connection between these men/symbols and cats!

If the agent were actually to see a cat a very similar process would ensue. The set or sets of sensor signals so generated would again be reduced to, say, a single man who would raise his flag upon "seeing" them, i.e. upon seeing a cat. This man is different from, and so far has no connection with, the other man. Notice again that there is no question of any special encodings and that the letters 'C', 'A' and 'T' are certainly not involved in the process. This should not be too surprising, afterall cats do not have the letters 'C', 'A' and 'T' stamped upon them in any way, and no amount of filtering or "image processing" could extract them from the vision of a cat!

What we have then, is a single set of input sensors giving two different patterns of signals in response to two different stimuli. Depending upon which of them is apparent, one of two internal symbols (men) signal this "fact". Notice that, if we were to select one of these men and trace out all the inputs to which he was connected (by following the arrows he had drawn), we would end up reproducing the sensor pattern to which he responds. In this way then, an agent may be said to have, or be able to generate, an internal (mental) image of, for example, the symbol 'C', however, it should be quite plain that this "pattern" is quite distinct from the external 'C' symbol. It is a "virtual" symbol produced by (corresponding to) a pattern of internal signal-symbols (men with flags raised).

The distinction between internal and external symbols is much more difficult to appreciate in the copy-storage scheme. Although ultimately we have exactly the same situation, there is a tendency to want, say, an external 'C' symbol, to somehow directly generate an internal ASCII 'C' code. But just as the letters 'C', 'A' and 'T' are not inherent to a cat or its image, so neither is the ASCII code in the form or appearance of a character.

Symbolic and Connectionist Paradigms Revisited

Let us now re-examine the symbolic and connectionist paradigms in the light of the preceding discussion. Notice first that questions of serial vs. parallel and analogue vs. digital, are irrelevant, they are simply alternative ímplementations (cf. communication techniques a and d respectively.) Also based on our definitions, both theories do indeed employ some form of symbols/signs to communicate and store information. In the case of the symbolic paradigm it should be obvious that storage is accomplished using the 'copy' method. Ostensibly the symbolic paradigm is based on a communication philosophy which demands encoding based on the fact that inputs from the environment are mutually exclusive (cf. communication techniques c and d). This raises the question of how to decide which sets of inputs might be suitable and how to guarantee that they actually are. There is no general solution to this.

In contrast the connectionist paradigm is quite clearly based on the 'link' storage scheme, although, in the case of artificial neural networks, it complicates this by adding the notion of a link "weight" which it uses to modulate the effect a particular input will have during the matching process. Deciding what weight to assign to a particular link is a difficult task, one which cannot be done in isolation because of its effect on the overall solution found by the matching process. Efficient automated 'learning' of appropriate link (interconnection) weights is a major research area in connectionism. Unfortunately, the whole idea of using weights is suspect, for they are the very essence of the discredited associationalist philosophy. The rationale behind the use of "weights" is that there is some fixed relation between each pair of concepts, what Smolensky referred to as the 'statistical connection' and Minsky called the 'importance' of a term. But this view is mistaken, there is no fixed degree of relation between concepts, rather it changes, often dramatically, with context. The situation is essentially that embodied in communication technique (b), see above, which requires identifying sets of input signals which will ALWAYS occur together. While it may be possible to arrange/show this in certain restricted environments it is clearly impossible in the general case.

Towards an Alternative Theory of Cognition

While the symbolic and neural network approaches are clearly unsatisfactory as general theories of cognition, the seeds of a possible solution are to be found in the above discussion. In particular the pure link-storage scheme, uncomplicated by the interconnection weights introduced by neural networks, seems to offer a suitable basis. As discussed so far, it is like a sort of filter, but of course filters are not very intelligent, so, if it is to offer a suitable basis for cognition, there must be something more. At the very least, it must explain both what and how concepts arise, and how external symbols can come to acquire meaning.

Assuming that particular 'men' respond to particular concepts as outlined previously, the latter question (how symbols can acquire meaning) is relatively straightforward to answer. What is needed is to relate - to form a link between - the man which responds to the letters 'CAT' and the man that responds to the image of a cat. This can be done in exactly the same fashion as before, that is, by introducing a new man with a flag, who has links to (draws arrows to) each of the other two men. Such a link-man could/would be established naturally as a result of the agent observing both a cat and the word 'cat' at (more or less) the same time. Put another way, the agent would store (using the link-storage method) a "situation" in which it saw both a cat and the printed word 'cat'. Now, as noted before, an intelligent agent must be able to check the present situation (set of input signals) against its stored knowledge, find the best match(s) and then "look" at what was the case in the immediate spatio-temporal vicinity of these prior situations. This knowledge constitutes the agents "expectations", based upon which it would select a suitable (intelligent) course of action. If now, our agent were to observe a cat, the fact that one of the situations it has stored in the past now includes one in which the word 'cat' was also observed, should led it to "bring the word to mind" too. The converse situation, whereby the word 'cat' was observed, would again cause the agent to recall the situation in which both word and object were seen, and hence bring the image of a cat to mind. In this way then, word and object are related such that external 'word' symbols gain appropriate "meaning".

This just leaves us with the difficult problem of finding appropriate "concepts" in the first place. We must somehow isolate the regularities in the environment, for these are the basis of useful concepts. But concepts are not always, or even just, a result of such regularities, nor are they individuals or simple statistical averages over all individuals. Moreover, what constitutes a particular concept, and even the set of useful concepts, may very well change over time. We have already seen that finding conjunctions of terms which always occur together is practically impossible (cf. communication technique b). The best we can do is to remember whatever we observe and match each new situation against this stored experience. If we arrange our link-storage scheme such that knowledge is stored in hierarchical form, that is, men can 'link' to inputs both directly and via other existing men, and we further insist that levels closer to the inputs are established before those higher up, then we should obtain natural groupings corresponding to interesting "regularities". It is not necessary (or even practical) to physically "restructure" this knowledge in order to identify new/different concepts since the matching mechanism already provides a way of identifying relevant groupings and their common expectations "on-the-fly". This, combined with the ability to relate word and object, would appear to provide a powerful enough basis for cognition.

Concluding Remarks

In this paper we have outlined a philosophy of symbols and used it to show why both major theories of cognition, those based on the symbolic and neural network paradigms, are ill-founded. The basis for a new, more viable, theory of cognition, was shown to arise naturally out of the critique developed here. The author's work on "inscriptors" [5,6] provides further support and details for these ideas.

References

[1] "Computer Science as Empirical Enquiry", Allen Newell and Herbert A. Simon, reprinted in The Philosophy of Artificial Intelligence, M.Boden, Oxford University Press, 1990. (ideas originally outlined in Newell, A. (1979) 'Physical Symbol Systems', Lecture at the La Jolla Conference on Cognitive Science. later published in Cognitive Science 4 (1980):135-83).

[2] "Connectionism and cognitive architectures: A critical analysis" Fodor J. and Pylyshyn Z., in Connections and Symbols (A special issue of the journal Cognition) (eds.) Pinker S. and Mehler J., Bradford Books, M.I.T. Press 1989, 1990.

[3] "Minds, Brains and Programs", John Searle, reprinted in The Philosophy of Artificial Intelligence, M.Boden, Oxford University Press, 1990.

[4] "A Theoretical Foundation for the Design and Implementation of Deterministic Digital Systems", David Davenport, Ph.d.. Thesis, University of Birmingham, England, 1980.

[5] "Inscriptors: Knowledge Representation for Cognition", David Davenport, in Proceedings International Symposium on Computer and Information Science-8, Istanbul, 1-3 Nov. 1993.

[6] "Artificial Minds: Fact or Fantasy?", David Davenport, Bilkent University Technical Report CIS9314.