Unarticulate creativity

T.D. Gedeon[1], **C.N. Quinn**[1] and **M.C.P. Wynter**[2]

[1] School of Computer Science and Engineering
[2] Law School
The University of New South Wales

Abstract

The general model of high-level thinking and problem solving is of a conscious, symbolic process. Recent advances in Artificial Intelligence modelling, such as connectionist approaches, are pointing out the role of low-level processing in cognition. Various phenomena such as verbal interference in performance support this view. Experience with neural network models demonstrate the reduction of dimensionality and loss of information in converting an internal representation to a symbolic form. We will discuss the models and evidence from a number of disciplines to present a case that an important component of thinking is conducted at a non-symbolic level.

1. Introduction

Why are we interested in the nature of thinking? Especially since we espouse a non-symbolic, unarticulable view of the process of thinking, why are we involved with an inherently symbolic process verbalising the unverbalisable?

The resolution of the dilemma is two-fold. Firstly, language is symbolic, but not in the same sense as mathematical or logical systems are, in that words have multiple, sometimes contradictory and mutable meanings. Secondly and most importantly, it is the only medium for complex communication. Yet, how often are we convinced by ? mere? intellectual argument? It is only after we have internalised the argument in some sense that we are convinced. Perhaps this explains the conference phenomenon. From a purely economic perspective, it is difficult to see the utility of transporting large numbers of fragile, temperamental and demanding objects (ie people), rather than just shipping the texts of their talks. Paper is compact, undemanding, and hardy allowing much economy in its transportation. Clearly, conferences are about more than the dissemination of pure symbolic information.

In the rest of this paper we will describe briefly the basics of neural network modelling, and briefly allude to the pervasive metaphor of symbolic thinking, before presenting counter-evidence from a number of disciplines. We will provide some suggestions towards a better model of thinking, however due to the very nature of our suggestions we can not by definition produce an exact model. We will close by providing a list of instructional implications / suggestions, and come to some conclusions about useful future directions.

2. Neural network modelling

This approach is premised on the use of a computational device which has limited information representation capabilities, and which carry out only very simple computations by exchanging simple messages along links: see Figure 1. Complex computations are done, not by a set of instructions operating on a representation of entire problems, but by massively parallel (actually or notionally) computations in interconnected networks of units or neurons.

Figure 1 A processing unit (neuron)

These connectionist devices ? neural networks ? are potentially usable where necessary and sufficient conditions for the application of concepts are absent. What is needed is reliable and representative data in the form of input-output sets ? that is, cases showing the occurrence of certain situational features or events as well as the categorised concepts which exist by virtue of these features. From such data the the system can make generalisations and potentially come to classify cases which are not part of the training set. In some connectionist models, the system is presented with data sets and a human trains the system by informing it in simple terms, such as whether one data set is or is not a member of a particular category. Clearly, there is no explicit symbolic representation of a concept. What we have is a distributed computational process which reproduces the behaviour of a human with respect to the concept.

The computational power of neural networks derives from simple local computations within individual elements, and their interactions with other processing elements, all operating in parallel. This is in contrast with the more familiar von Neumann computer architecture we find in modern digital computers, with their large, fast central processing unit which performs large numbers of calculations in sequence. It is worth pointing out at this stage that while hardware to implement the requirements of truly parallel neural networks exist, and are becoming cheaper, most work and utilisation of neural network methods still use simulations of the neural network on von Neumann machines. This points out two important aspects. Firstly, that a new paradigm enables work to be done which could not be readily done before even though the same underlying computer is still being used. Secondly, that we can simulate these networks on our von Neumann architectures means that we have not discovered something magical, just a very useful new way of looking at and solving some problems.

We will here briefly introduce the most commonly used neural network model, being the feed-forward network trained by back-propagation, and later mention a simpler neural network model when we make a point about emergent properties.

Feed-forward networks usually contains three layers, as shown in Figure 2. More layers are not really necessary because for every network with more than three layers it is possible to construct a network with only three layers to carry out the same processing.

Figure 1 Feed-forward network

Signals arrive from the outside world at the input layer, are processed there, and the results passed on to the next layer - the hidden layer. The hidden layer is so-called because it is connected neither to inputs nor outputs; hence, it is not ?visible? to the outside world (Touretzky and Pomerleau, 1989). The units in the hidden layer and those in the output layer, similarly process their respective incoming signals and send results forth. The collective result of the units in the output layer - shown as signals on the outputs - may be considered to be the system?s purported answer or solution to the query or problem posed by the signals sent from the outside world via the inputs.

During training or learning, a number of test cases are presented in succession to the network?s inputs. The respective output result is compared with what is considered to be the correct answer or solution; the difference is denoted in an error measure. Where the error consisted of a wrong activation of a unit in the output layer, the weights of the links coming into that unit are modified to depress or increase the activation - as the case may be - to reduce the error. These modifications of the weights of the links between the hidden and output layers has implications for the ?correctness? of the weights of the links between the previous two layers. This is because one can think of the error at the output having arisen due to a cumulation of errors at each step of the propagation of signals. This process of alteration of weights in accordance with an error measure is called ? back propagation.?

When a solution is correct, we can say that the links and their weights are such the network possesses the knowledge as to the relationships between the relevant inputs and outputs. The implication is that the superposed knowledge in respect of the test cases used for training or learning is collectively adequate for the purpose of finding solutions to new problems in the same domain.

This model is probably the most popular in applications involving generalisation, and account for an estimated 70%-80% of networks in use. It has been applied to areas such as mortgage loan authorisation and bond rating. These areas have a large number of variables which may influence the outcome of analysis and assessment of risk, yet it is rarely clear which factors are indeed relevant (Mital and Gedeon, 1992). Feed-forward networks have achieved significant improvements over statistical techniques (Dutta and Sekhar, 1988). This suggests that the sub-symbolic processing of the networks has been able to discover at least some of the significant features which are not taken into account in the statistical models.

3. Thinking Symbolically

That thinking is a symbolic process has a long philosophical derivation from Platonic absolutes; that there is a deeper underlying structure to be discovered from which we can understand thinking.

Traditional AI is based on a similar view. The physical symbol hypothesis of Newell (1980, Newell and Simon, 1981) considers that knowledge is symbols of reality and the relations between symbols, and merely need appropriate logical manipulation to produce thought.

In the realm of physics, for example, we have had to give up on such absolutes. Atoms are not indivisible; from quantum mechanics we know that the universe is inherently fuzzy at the smallest scale.

4. Counter-evidence

Here we will discuss evidence from a number of disciplines. Given our premise of the non-symbolic nature of true higher level cognition, we unashamedly start with a number of persuasive colloquial arguments in support of our stance.

4..1. Colloquial evidence

There is a great deal of colloquial evidence for a non-symbolic view of thinking. We have all been in the situation that a solution to a complex problem just springs to mind. The traditional ? eureka? is a good example. We are often exhorted to ? sleep? on complex and seeming unsolvable problems. We try to think laterally, and so on.

The concept of human thought as non-symbolic has appeared a number of times in the fantastical literature known as science fiction - the literature of the scientific method. The clearest expression of the symbolic vs. non-symbolic dichotomy is in Shaw (1989). The antagonist in the key scene is a telepathic alien with definitely symbolic thought processes, as we know from our view of its introspections, particularly in contrast with its awareness of the thought processes of the protagonist. The protagonist? s thought processes are only telepathically readable when conscious. The protagonist surprises himself with a useful verbal utterance. The alien is flabbergasted that a non-symbolic process has produced a meaningful suggestion.

The act of surprise at one? s own words is a common phenomenon, experienced in highly focused and demanding situations such as the teaching of difficult concepts. The ability to be surprised at our own intelligent actions indicates that not all intelligence is conscious.

4..2. Extraction of rules from a trained neural network

This area has received attention, because it seems that an ideal combination of the functionalities of neural networks and knowledge systems would be to train a neural network using examples, somehow extract a set of rules, and then use the rules to build a rule based knowledge system. In this way, some of the perceived deficiencies of neural networks could be corrected. A prime example is the lack of provision of explanations of the conclusions reached by a neural network. Also, the extensive technology for using, testing, and evaluating knowledge systems would become available. This scenario is particularly attractive for domains where very few experts are available, or the expertise is particularly difficult to formalise.

The problem with extracting rules from a neural network is mainly a matter of how not to extract rules. That is, the entire network could trivially be expressed in terms of a very large number of rules which would effectively simulate the processing of the neural network. The problem is to extract only a few rules, which somehow capture the important properties of the neural network solution. While arguably it is not possible to completely capture the knowledge inherent in the sub-symbolic distributed representation used by the neural network, in practice some progress has been made. The best of a number of recent related approaches is that of fuzzy rule extraction (Enbutsu, Baba and Hara, 1991).

A fuzzy rule in this case is expressed in terms of high, medium and low values of the input and output variables. This simplifies the complexity of the task of rule extraction as there is no requirement to include more precise values. These values would otherwise have to be derived from the inputs during training, and matched with the relevant rule being extracted. The separation of values into high, medium and low categories is done by replicating the units in the input and output layers threefold, with a single unit now representing a boolean value indicating whether its variable is in the specific range.

After training, the network contains an internal representation of the knowledge it has acquired in the form of weights. The acquired weights are evaluated using (relative) causal indices. A causal index is a measure of the causal connection between a unit in the output layer and a unit in the output layer.

The causal indices for units are calculated on a presentation of a cluster of training patterns from a specific classification. This clustering further simplifies the task allowing rules for separate classifications to be extracted separately. The task is further simplified by only examining the causal index for selected factors. Causal relationships which are insignificant are discarded, and a set of rules are extracted in the form of *if-then* rules, using conjunction and disjunctions.

In the above process, some information is discarded at each step. Only a number of factors are considered, and rules are only extracted for clusters of related patterns from a single classification. This has some interesting parallels with the difficulties encountered in knowledge elicitation from experts. That is, the the rules we extract will be quite different depending on the classification the cluster of patterns is drawn from. This classification and the cluster of patterns is the *context* for which the extracted rules are meaningful.

4..3. Human information retrieval

Investigators have generally found that the cues most effective in accessing a memory trace at retrieval are those most similar to its encoding context. This evidence suggests that a person possesses a tremendous amount of loosely organised knowledge for a category in long-term memory, however only a very small subset is ever active on a given occasion to represent the category in working memory (Barsalou, 1989).

A similar structure has been used (Gedeon et al, 1992) in a simple interactive neural network model for information retrieval in a legal support activation. The network automatically discovers synonyms for query items based on their context in documents. The network recognises query concepts by the pattern of activation of the neurons for conceptual information retrieval.

4..4. Better left unsaid

Schooler and Engstler-Schooler (1990) have documented the deleterious effect of reflection on recall. The verbal overshadowing effect is the impairment of subsequent recognition of the original stimulus by verbalisation of hard to describe stimuli. In their study, subjects who described a face after presentation were significantly worse on a recognition test than subjects who were not asked to describe the face. This has profound social implications in a number of domains such as witness testimony in trial procedures, and identification of suspects by Police identi-kit methods.

Further studies have suggested that the verbal overshadowing effect may generalise to broader cognitive processes. A study by Wilson and Schooler (1991) have shown that verbalisation can affect the quality of affective decisions, and a study by Sieck (1991) has demonstrated this effect on the judgement of similarities.

Clearly, verbal description is symbolic in nature. The very nature of the face task as ? hard to describe? is suggestive. If the information stored about the face is symbolic in nature, it is surprising that so much information is lost during the translation to a verbal (symbolic) language. If, conversely, the information about the task is stored in a non-symbolic fashion, the loss of information in the creation of a symbolic description is natural.

4..5. (symbolic) Knowledge elicitation from experts

The Platonic view that there is a deeper underlying structure to knowledge is problematic. The earliest refutation of this view was by Zhuang Zi, a philosopher of ancient China, as depicted by Tsai (1989) in Figure 3.

Figure 1 Zhuang Zi on knowledge acquisition

There is a dichotomy, or lack of relation between what we do and what we say we do (Compton and Jansen, 1990). The physical symbol hypothesis has become so pervasive, that the conventional explanation for the difficulties in knowledge acquisition is that the pieces of basic knowledge are assumed and combined too quickly to be readily described. Based on the experience with the Garvan Expert System, and the few other systems in long term use, Compton and Jansen (1990) suggest that knowledge in the form of logical rules as acquired from experts are essentially a justification of why they are right, and not the reasons why they reached the conclusion. Knowledge is thus strictly contextual in nature and relate more to the demands of the situation rather than to information.

5. Thinking non-symbolically

In a deliberately disjointed fashion, here are some observations on non-symbolic thinking.

Symbols are too useful to throw away completely, perhaps what we need is a more dynamic and changing definition of a symbol, rather than the dead symbols of logic and mathematics. A live, or active symbol can be the relationship between a collective phenomenon and the small scale events it is made up of, such as a hypothetical intelligent ant hive (Hofstadter, 1979). Such collective phenomena inextricably linked with their context, which ties in well with current trends towards ? situated? cognition, situated problem solving and knowledge acquisition.

What is creativity but a highly desirable expression of deeper processes. Creativity is in particular considered to be a high level function. Conscious thought is also considered high level. The two are thus often conflated, as illustrated in Figure 4.

Figure 1 Zen and symbolic thought

Zen is a Mahayana Buddhist movement dating from the 6th Century AD in China. Very briefly, the emphasis of this movement is on achieving enlightenment by direct intuitive insights, as opposed to explicit teaching. It seems we may not have discovered a new model after all, but are involved in the rediscovery of an older model. Zen is certainly has similarities to the approach we suggest. Nevertheless, we prefer a more active approach than is found in traditional Zen, with its contemplative and abnegatory attitudes towards thought.

6. Instructional Implications

All of the implications are straightforward, but we need to discover these for ourselves, individually. In this section we provide some rough guidelines.

- Teach by doing, not (just) talking.

Many of us do this anyway. Chemistry is traditionally taught using a laboratory environment, yet cynics suggest that all students learn is the ability to fudge results, and copy lab reports as the same experiments are repeated year after year. Yet this situation is not too dissimilar from the paucity of information available during children? s acquisition of language.

Talking is not wholly symbolic in our view, all human interactions are dramatic occurrences, every text is a performance text with the audience receiving a very message strained through the colour of their personal glasses.

- Doing is situated, talking about doing is better than just talking.

We communicate using word symbols, thus they have some use. Certainly, in the realm of symbolic Artificial Intelligence, there are useful approaches to try to solve the problems posed by what we have postulated as the non-symbolic nature of thought. In the field of expert system development, the recognition of the situatedness of real knowledge has resulted in the use of ripple-down rules which link rules with the contexts causing the enhancement of the rule base (Compton and Jansen, 1990).

This has similarities to the idea of teaching by doing exercises which produce different results and so on. This use of examples is also quite similar to the use of examples in training neural networks. In training neural networks, the selection of training patterns to use, and the pre-processing of the raw information are all important to the successful development of a practical application. This is again not dissimilar to the careful selection of exercises and information to present by teachers for appropriate human learning to take place as distinct from the memorisation of rules.

- Teach in a well-defined context.

As mentioned earlier, the most successful cues for retrieving a memory trace are those similar to its encoding context. This has two implications, firstly that we should try to teach in an active and interactive fashion where there is some correspondence to the usage of the information learnt rather than in a didactic class-room or lecture fashion.

■ Practice skills across a number of contexts.

Need to practice thinking skills across domains to increase the likelihood of successful retrieval of cues and so on. This can be in the form of actual presence or in terms of visualisation and role playing. The benefit of computer technology in such interactive teaching is thus profound (Quinn, 1991, 1992). The second implication is that we should enable students to use techniques such as visualisation and role playing to enhance their own retrieval of information.

■ Overload symbolic reasoning.

A Zen koan is a paradox used in teaching (meditation), a bit of apparent nonsense that forces one to abandon dependence on reason and break through to sudden intuitive understanding (Salajan, 1982). It is very hard to stop thinking of elephants after such a request has been made, it is even harder to avoid thinking about thinking. A better way is the koan method, overloading logical reasoning by concentrating on a paradoxical situation.

7. Conclusions

Is creative thinking conscious? Probably not.

We have presented a largely symbolic, and partially non-symbolic (visual) argument in favour of our contention that an important component of particularly higher level thinking such as creativity is conducted at a non-symbolic level.

We conclude that we need some articulation management strategies - to reduce our dependence and focus on articulation with its largely symbolic roots.

References

Barsalou, LW ? Intraconcept similarity and its implications for interconcept similarity,? in Vosniadou, S and Ortony, A (eds) *Similarity and analogical reasoning*, Cambridge University Press, 1989.

Compton, P and Jansen, R ? A philosophical basis for knowledge acquisition,? *Knowledge Acquisition*, vol. 2, pp. 241-257, 1990.

Dutta, S and Shekhar, S ? Bond rating: a non-conservative application of neural networks? , *Proc. International Joint Conference on Neural Networks*, vol. II, pp. 443-450, 1988.

Enbutsu, I, Baba, K and Hara, N ? Fuzzy Rule Extraction from a Multilayered Neural Network,? *Proc. International Joint Conference on Neural Networks*, vol. II, pp. 461-466, Seattle, 1991.

Gedeon, TD, Johnson, L Mital, V ? Neural Networks and Legal Reasoning,? in Mital, V, Johnson, L, *Practice Support Systems in Law: Applying Strategic Technologies*, Chapman & Hall, 1992.

Hofstadter, D *G? del, Escher, Bach: an Eternal Golden Braid*, Basic Books, 1979.

Mital, V and Gedeon, TD *Neural Experts and Information Systems in Finance*, Chapman & Hall, (in press) 1992.

Newell, A ? Physical Symbol Systems,? *Cognitive Science*, vol. 4, pp. 135-183, 1980.

Newell, A and Simon, ? Computer Science as Empirical Enquiry: Symbols and Search,? in Haugeland, J (ed) *Mind Design*, MIT Press, Cambridge MA, 1981.

Quinn, CN ? Teching Cognitive Skills with Computers,? *Proc. of East-West Conf. on Emerging Comp. Tech. in Education*, Moscow, 1992.

Quinn, CN ? Computerized Cognitive Research: A HyperCard Adventure Game,? *Behavior Research Methods, Instruments, & Computers*, vol. 23 (2), pp. 237-246, 1991.

Salajan, I *Zen Comics II*, Tuttle Books, 1982.

Shaw, B *The Wooden Spaceships*, 1989.

Schooler, JW and Engstler-Schooler, TY ? Verbal Overshadowing of Visual Memories: Some things are better left unsaid,? *Cognitive Psychology*, vol. 22, pp. 36-71, 1990.

Sieck, W ? Verbal Overshadowing Effects on the Judgement of Analogy,? Unpublished thesis, University of Pittburgh, 1991.

Touretzky, DS and Pomerleau, DA ? What? s hidden in the hidden layers,? *Byte*, August, pp. 227-233, 1989.

Tsai, CC *The Sayings of Zhuang Zi*, Asiapac, Singapore, 1989.

Wilson, TD and Schooler, JW ? Effects of Self-Reflection on Preferences and Decisions: Who agrees with the experts?? Unpublished article, University of Virginia, 1991.