3th International Conference on Public Communication of Science and Technology (PCST), Montreal,Canada,10-13 April,1994

WHY NEURAL NETWORKS ARE NOT BICYCLES: SOME PROBLEMS WITH THE "SCOT" APPROACH TO CERTAIN TECHNOLOGIES

Christopher Carne,
Centre For Research into Innovation, Culture and Technology,
Brunel University, Uxbridge, Middlesex, UB8 3PH
England

Introduction

The Social Construction of Technology (SCOT) (Bijker and Pinch 1987) approach is a widely known and influential framework within STS. Woolgar (1991) has argued that this (and related) approaches are ultimately dependant on "conventions of realist discourse" and that such an approach fails because it has not adequately addressed this dependency. In this paper I wish to show how the reliance on such realist discourse is foregrounded when we come to study certain technologies and conversely how such a reliance seems unproblematic when studying other technologies. I shall argue that not only does the failure to address such questions shows a compromise of the "radical potential" of SSK, it also conceals certain empirical questions that are central to the study of technology.

To pursue this aim I shall begin by outlining how the SCOT framework would approach the development of two different technologies. The first being the bicycle, as is discussed by Bijker and Pinch (1987), the second being the neural network. I then show that while such an account appears adequate, it rests on certain analytic assumptions. Assumptions that are themselves part of the social construction of technology and, as such, open to interrogation. In conclusion I will suggest that Woolgar's (Woolgar 1991, Cooper and Woolgar 1993) notion of "technology as text" provides us with a better framework for considering the development of both neural networks and bicycles.

What is a Neural Network

While most of us know what a bicycle is, considerably fewer people are familiar with a neural network and it is probably worth digressing in order to give some idea of the sort of artefact under discussion. Neural computational (aka connectionist) methods have become increasingly important in the last ten years or so, finding applications in many different areas from natural language processing to image recognition tasks. They have also been claimed by cognitive scientists as providing an alternative method for modelling cognition to that which has traditionally been used in AI. Perhaps the most succint definition is given by Aleksander and Morton (1991).

> **Definition:** Neural Computing is the study of networks of adaptable nodes which through a process of learning from task examples, store experiential knowledge and make it available for use. (p. 1)

This account gives the key features of most neural networks. The simplest form of a neural network device is Rosenblatt's "Perceptron" (Rosenblatt 1959, 1962). Rosenblatt used a simple adaptable node based on McCulloch and Pitts' "formal neuron" (McCulloch and Pitts, 1943), provided a simple training rule and showed how it could store certain knowledge. Although most modern neural networks have more complicated architectures and more sophisticated learning rules they have many features in common with the perceptron. The important feature to note here is that the perceptron adapts by changing the weights associated with each input in accordance with a learning rule. So, if we wish to train a network to store a set of four patterns, e.g., the Boolean AND, we will present each pattern in turn, and update the weights until the correct output is achieved in all cases. Then the network can be used as an AND gate. It has stored the AND function and makes it available for use.

A SCOT Approach to Neural Networks

In developing their model for the development of the bicycle Bijker and Pinch define three basic terms, relevant social groups (henceforth RSGs), interpretive flexibility and closure. I shall outline each of these in turn.

Bijker and Pinch define RSGs as follows:

> The phrase is used to denote institutions and organizations (such as the military or some specific industrial company), as well as organized or

> unorganized groups of individuals. The key requirement is that a certain social group share the same set of meanings attached to an artefact. (p. 30)

For example, some RSGs for bicycles would be cycle designers, manufacturers and users. Within these groups might be sub-groups such as women cyclists, sports cyclists, touring cyclists, etc. One of the key roles of these RSGs is in the definitions of "problems";

> a problem is defined as such only when there is a social group for which it constitutes a 'problem'. (p. 30)

For a sports cyclist, the major problem might be one of speed, for a touring cyclist, one of safety and so on. Of course, for each problem there may be may solutions (both social and technical). The relationship between artifacts, RSGs, problems and solutions can be seen in Figure 2.

It is clearly possibly, given enough ingenuity (and a decent graphics package) to produce such a representation of the RSGs involved in connectionism. There are a number of clearly identifiable RSGs (Figure 3) and many definitions of problems and many design solutions. To examine just one aspect of this we may consider the problem of "hard learning". In the well-known book "Perceptrons" (Minsky and Papert, 1968), which may be taken as a representative of the "Good Old Fashioned AI" (GOFAI) RSG, it was suggested that there were certain types of problems, so-called "hard learning problems", that caused fundamental difficulties for neural nets. An example of this was the inability of the perceptron to learn the Boolean exclusive OR (XOR) function as the function is not "linerarly seperable" (Figure 4).

The solution to this problem was seen by the pro-neural network RSG as solvable by "dimensional expansion" by using multi-layer perceptrons (MLPs), thus making the XOR function linearly separable. Figure 6 shows a simple MLP. The MLP approach was then seen as problematic by several RSGs, both pro and anti-connectionism. Problems such as how to train the MLP became important, the so-called "structural credit assignment problem". There are now many more or less stabilised methods for overcoming this problem. The first of these (perhaps) and probably the best known is "error back propagation" attributed to Werbos (1974) but more widely publicised by Rumelhart and McClelland (1986). Other groups involved in connection believed the MLP family of neural networks were overly complex, and proposed solutions such as pre-coding of input data with much "simpler" neural nets, this problematisation and solution is associated with the RSG of those working on "weightless" neural networks. So we can see, even in such a sketchy account of neural networks that the idea of RSGs seems to be quite useful.

The second of Bijker and Pinch's concerns is that of "interpretive flexibility". In their definition of this term it refers to the different design possibilities as well as the different interpretations that an artifact might have. We have already seen a good example of this in the solution to the XOR problem. MLPs being seen as a good solution by one group of scientists, and as needlessly complex by another. Likewise, the design solution to the problem has flexibility, the weighted and the weightless approach both being presented as equally solving the problem.

The final concept that Bijker and Pinch employ is the notion of stabilisation or closure of technical change, that is, the disappearance of problems. They suggest that this occurs either by rhetorical means or by a redefinition of the problem. The first of these involves convincing the RSGs that a problem has been solved and it is not the case that this has been achieved in the case of hard learning. The second involves changing the problem to fit the solution. So the "hard learning problem" becomes "the structural credit assignment problem". Not all the RSGs are satisfied that solving "the structural credit assignment problem" entails the solution of all hard learning problems and there is not much sign anything more than local closure at the moment.

Well, We All Know What a Bicycle is, But What's a Neural Network?

I wish to move on to suggest that this account of the development is somewhat problematic.

Woolgar (1991) has suggested that SSK is dependent on "conventions of realist discourse"; moreover he claims that;

> [T]he deconstruction of the selected technologies depends upon what pass (in the curse of the argument) as definitive versions of the capacities and

> attributes of these technologies. This occurs despite the axiomatic appeal to the interpretive flexibility of the character and attributes of technology and the disavowal of technological determinism. (p. 35)

In the brief outline of what a SCOT approach to the development of neural networks might look like we find just these dependencies. It has been suggested that there are a class of artefacts that have been identified as multi-layer perceptrons and that the development of this technology has been subject to the interaction of various RSGs providing their interpretations and attempting to impose closure. In this simple account we have already assumed quite a lot about the nature of the artefact under consideration. The particular assumption that I wish to examine here is that such artefacts exist, external to the representational practices that surround them. Moreover, I shall suggest that the very notion of an artefact is a social achievement and should not be taken as one of the definitive attributes of a given technology.

Most people have a reasonable idea of what constitutes a bicycle, but what is a neural network? If we considfer one particular definition we can get some idea of the potential diversity of artefacts that might be called neural networks.

This schema (Figure 5) is drawn from Aleksander (1989). We can consider such a schema as an attempt to produce a "recogniser" for neural networks. Thus, if we accept Aleksander's recogniser we have a sufficient, if not a necessary way of telling whether any given artefact should be considered as a neural network . So we can see that the representation given by Bechtel and Abrahamsen (1991) (Figure 6) should indeed be considered as a neural network as it fulfils the criteria demanded by Aleksander's recogniser. Likewise, I was able to give Aleksander and Morton's definition of what a neural network was and to show that the perceptron was a neural network.

The question we now come across is the status of this generalisation. Work by Latour(1988) and Cartwright (1983) has problematised the notion of generalisation, but even staying within the SCOT approach we find a major problem. Any account of what a neural network is bound to be a interpretation. This being the case what "metric" can we use to ascertain whether the various RSGs are interpreting the same artifact? Clearly, in the case of neural networks, the boundedness of the object is no longer obvious. “Boundary work”, such as the production of general formalisations must be done in order to establish the bounded, or “real”, artefact.

In order to clarify this point, imagine that we knew as much about bicycles as most people do about neural nets. We would have no way of knowing that the penny-farthing and the “singer extraordinary” were both examples of the class of artefacts designated as bicycles. In addition, on being told that they were both bicycles we would have no way of knowing what the significant aspects which allowed this classification were without asking members of the RSGs.

We can see the process of attempting to construct the identity of an artefact in the case of the perceptron. The story has been told of how the single-layer perceptron could not solve hard learning problems but that a multi-layer perceptron could once the structural credit assignment problem was solved. If we dig a little deeper into this story we find that an *a priori* distinction between the single and multi-layer perceptrons is impossible to maintain. A common way of establishing the failure of any given attempt to solve the structural credit assignment problem was to show that the resulting network was formally (and thus actually) a single-layer perceptron and thus open to the Minsky and Papert critique. This achievement of identity between apparently diverse artefacts being achieved by formal mathematical demonstrations. Demonstrations that are of varying degrees of complexity but all depending on the pre-constructed of mathematical knowledge with its associated proof mechanisms and conceptions of identity and difference.

Again, we see the problem with assuming that the artefact is a pre-existing given open to differential interpretation. It may be of no great moment that we can demonstrate that a neural network is not a bicycle or that a penny-farthing and a modern racing bicycle are both variations of the same artefact, yet it was a key factor in the development of neural computing to be able to show that a given neural network was/was not a single layer perceptron.

So, far from there being an artifact that is then given meaning by RSGs, the only way that we can identify an artefact is through the meanings attributed by the RSGs. It is only ever possible to refer in general to “the bicycle” because there is a general RSG (which includes Bijker and Pinch) which has achieved closure on the classification metrics for what constitutes a bicycle. In the case of neural networks, the classification metrics are sufficiently complex and far less stabilised, so these features force us to attend to the metrics themselves, whereas when considering as more mundane technology the metrics become hidden.

So, Show Me a Neural Network

In discussing the ontology of neural networks, we have so far neglected to consider the even more complex question of representation of neural networks. I now wish to show that interpretive flexibility applies as much to the representations of neural nets as it does to the neural net as artifact. Moreover, it is the constitutive nature of representations that account for the problems we have in considering neural networks as pre-existing artefact.

There are a wide range of representation strategies, both formal and informal, used in the field. The examples I have here show several common varieties, pictoral, mathematical, and computer code. Figures 5 and 6 from Aleksander (1989) and Bechtel and Abrahamsen (1991) we've seen already. Figure 7, Churchland (1989), gives a pictorial description of the behaviour of a neural network. Rummelhart and MacClelland (1986) (Figure 8) give a representation that relates the neural network to biological neural systems. Figure 9, Caianiello (1989), is part of a formal mathematical decription. Figure 10 is a partial listing of the source code of a neural network program supplied with Rummelhart and McClelland (1988). Figure 1 is my representation of the perceptron.

How then are we (as analysts) to cope with such a huge diversity of representations that are dealt with as a matter of routine by participants? I suggest that we must consider these representations as artifacts and that they are as much subject to interpretive flexibility as physical objects. Of course, we encounter the same problem as we had when trying to use the concept of bounded objects previously. Certain types of representation will be held to be equivalent, and some different. For instance, a common assumption amongst most (if not all) RSGs in the mathematical field is that algebraic and graphical representations are equivalent. Of course, as a participant in the field, we can draw upon the discursive assumptions that assign difference and similarity to mathematical objects, but as an analyst we should attend to the ways in which the discourse functions in scientific practice.

If we claim that representations are interpretively flexible and are characterised by RSGs, we, as sociologists, should also attend to the ways in which we represent. Sociologists are, after all, RSGs themselves. Indeed, by attending to the interpretations that we make ourselves, whether it's the identification of RSGs, accounting for science, assigning functions to representations, characterising the work of other sociologists, presenting versions of "the technology", etc., we gain insight into the general question of interpretive flexibility with regard to technology and its representation.

The Reflexive Option: Technology as Text

The position that I am arguing towards with the above example is close to that of Woolgar (1989) in presenting "the reflexive version of the technology of text metaphor". He suggests that this approach includes the question:

> How is the reality of the technology itself created, described sustained, and, in particular, how do the effects and the capabilities technology relate to the effects and capabilities of other entities in the text in which they are inscribed. (p. 42-43)

In addition, Woolgar suggests that we should be wary of versions of constructivism in which

> [T]he essence of technology is presumed unchanged; it is merely being called, rendered, described as something different. (p. 42)

It may well be that the deeply reflexive nature of representational practice in neural network research makes it a "soft case" by which to make arguments for an argument such as Woolgar's. It is clearly the case that representations of neural networks are the main focus of practice. Various representations are produced and manipulated. Claims are made and verified with regard to representations. It can be claimed that the practice of neural network research is the practice of building and manipulating representations. The separation between object and representation is particularly difficult to maintain in this case. In this particular case, the technology is text, no metaphor is involved. Given this it is inappropriate for the analyst to attempt to produce representations of the technology that are somehow natural or neutral. We are either making representations that serve our own RSGs or replicating representations that are constructed by RSGs within the neural network community. While this is unavoidable it should be foregrounded in the analysis rather that ignored or hidden. Realism is a complex achievement of the participants in the field despite the reflexive nature of their practice. The job of the analyst should be to mark this achievement and follow its dynamics, both in scientific practice and, reflexively, in our own practice.

Conclusion

So what, after all, are neural networks? The question, given a more radical interpretation of interpretive flexibility makes the question somewhat meaningless. We have seen that they are a sequence of mathematical expressions, computer programmes, a sequence of tuples, a model of human cognition, and so on. They are also a device for arguing against the SCOT approach, a way of problematising of the sort of essentialist position that allows us to ask the question "What is a neural network?" and something by which the need for a reflexive approach to STS can be argued.

Now the task is to go back through this paper and remove all the essentialist claims about Bijker and Pinch, RSGs, reflexive sociology, connectionism and, so forth. Or at least to note the processes that allow such claims to be made.

Bibliography

Aleksander, I. (1989). "The Logic of Connectionist Systems" in Aleksander (Ed) 1989, p. 133-156.

Aleksander, I. (ed) (1989). *Neural Computing Architectures: The Design of Brain-like Machines*, North Oxford: London.

Aleksander, I. and Morton, H. (1991). *An Introduction to Neural Computing* Chapman and Hall: London.

Bechtel, W. and Abrahamsen, A. (1991). *Connectionism and the Mind*, Blackwell: Oxford.

Bijker, W. and Pinch, T. (1987). "The Social Construction of Facts and Artifacts" in Bijker, W., Hughes, T., and Pinch, T. (Eds) *The Social Construction of Technological Systems*, p. 17-51, MIT Press: Cambridge, Mass.

Caianeillo, E. (1989). A Theory of Neural Networks' in Aleksander, I. (Ed) 1988 p. 8-26.

Cartwright, N. (1983). *How The Laws of Physics Lie,* Clarendon Press: Oxford.

Churchland, P.M. (1989) *A Neurocomputational Perspective*, MIT Press: Cambridge, Mass.

Cooper, G. and Woolgar, S. (1993). "Software Quality as Community Performance" CRICT Discussion Paper No. 37, Brunel University, Uxbridge, Middx.

Latour, B. (1988). "The Politics of Explanation: An Alternative" in Woolgar, S (Ed) *Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge*, p. 155-177 Sage: London

McCulloch, W. and Pitts, W. (1943). "A Logical Calculus of the Ideas Immanent in Nervous Activity" in *Bulletin of Mathematical Biophysics*, 5, p. 115-133

Rosenblatt, F. (1959). "Two Theorems of Statistical Seperability in the Perceptron" In *Mechanisation of Thought Processes: Proceeding of a Symposium Held at the National Physical Laboratory*, Nov 1958, Vol. 1, p. 421-456, HM Stationary Office: London.

Rosenblatt, F. (1962). *Principles of Neurodynamics* Spartan: New York.

Rummelhart, D. and McClelland, J. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, *Foundations*, MIT Press: Cambridge, Mass.

Rummelhart, D. and McClelland, J. (1988). *Explorations in Parallel Distributed Processing*, MIT Press: Cambridge, Mass.

Werbos, P (1974). "Beyond Regression: New Tools for prediction and Analysis in the Behavioural Sciences", PhD. thesis, Harvard University, 1974.

Woolgar, S (1988). *Science: The Very Idea*, Ellis Horwood: Chichester.

Woolgar, S (1991). "The Turn to Technology in the Social Studies of Science" in *Science, Technology and Human Values*, Vol. 16, No. 1, Winter 1991, p. 20-50.

Nota

We cannot read the Figures in the text. Please, see the author.