The Pragmatic Roots of Context

Bruce Edmonds

Centre for Policy Modelling,
Manchester Metropolitan University
http://www.cpm.mmu.ac.uk/~bruce

Abstract

When modelling complex systems one can not include all the causal factors, but one has to settle for partial models. This is alright if the factors left out are *either* so constant that they can be ignored *or* one is able to recognise the circumstances when they will be such that the partial model applies. The transference of knowledge from the point of application to the point of learning utilises a *combination* of recognition and inference – a simple model of the important features is learnt and later situations where inferences can be drawn from the model are *recognised*. Context is an abstraction of the collection of background features that are later recognised. Different heuristics for recognition and model formulation will be effective for different learning tasks. Each of these will lead to a different type of context.

Given this, there are (at least) two ways of modelling context: one can either attempt to investigate the contexts that arise out of the heuristics that a particular agent actually applies (the 'internal' approach); or (if this is feasible) one can attempt to model context using the external source of regularity that the heuristics exploit. There are also two basic methodologies for the investigation of context: a top-down (or 'foundationalist') approach where one tries to lay down general, *a priori* principles and a bottom-up (or 'scientific') approach where one can try and find what sorts of context arise by experiment and simulation.

A simulation is exhibited which is designed to illustrate the practicality of the bottom-up approach in elucidating the sorts of internal context that arise in an artificial agent which is attempting to learn simple models of a complex environment. It ends with a plea for the cooperation of the AI and Machine Learning communities as *both* learning and inference is needed if context is to make complete sense.

Keywords: transferrence, learning, inference, context, heuristic, pragmatism, modelling, methodology

1. Introduction

Frequently at workshops and conferences on context, one finds that the emphasis is on drawing distinctions between different types of context and illustrating how little each type has to do with the others[1]. If this trend continues it will quickly become impossible to use the term "context" at all. Now it is certainly the case that naively conflating different usages of the term can cause confusion, but I wish to claim that there is a good reason that we use the same term for these different entities. The reason, I claim, is that context arises from a study of the *pragmatics* of learning and applying knowledge. These roots of context explain why and how the different types of context and approaches to studying them arise. This account centers on the *transference* of knowledge between learning and application. If this is the case, then accounts of context which capture either *only* context-dependent learning or *only* context-dependent inference will be inadequate.

1. This was the case at ECCS'97 and also reflects Pat Hayes' summation of the two AAAI Fall symposia on the subject.

This paper is structured as follows: section 2 is about the causal structure of complex systems showing the inevitability of the selection of important factors in any model we construct; this motivates the account of context as an abstraction of the features that are not explicitly included in the model but used in the recognition of its applicability in section 3; section 4 relates how the choice between concentrating on the actual heuristics used by an agent and on the external regularities that they exploit lead to the familiar 'internal' and 'external' approaches to context; the account also explains why different sources of commonality will result in the different types of context encountered in different fields (section 5); section 6 traces two possible methodologies for the investigation of context ('top-down' and 'bottom-up') noting the present bias towards 'top-down' studies; in order to show the viability of 'bottom-up' studies taking the 'internal' approach to modelling context a simulation study is described in section 7; I conclude in section 8.

2. Causal Structure

In any but the simplest (e.g. linear) systems, effects can have a great many causes (when modelled in absolute terms). In most systems, the web of causation is so dense that the number of factors that could be included in a model of an event is limited only by the resources we put into it. This is what has been called "causal spread" by Wheeler and Clarke in [19]. It has led some philosophers to argue that many (i.e. non-statistical) notions of causality are only meaningful for simple systems (e.g. [18]). This causal spread makes the *complete* modelling of an event impractical – we are forced to concentrate only on a small subset of possible factors. Of course, this omission of factors in our models is only effective if *either* they are so reliable that their omission is unimportant for practical purposes *or* if we are able to recognise when our restricted model is likely to be applicable. In either case we will not attempt to use the model when it is inapplicable.

Formalisations of causality always involve assumptions about the set of possible factors. Usually they merely present a test which can be used to reject the hypothesis that a given factor or variable is causally irrelevant. The strongest formulation I have found is that of Pearl [15]. He presents an algorithm for finding all the factors that *are* causes, but under the assumption that *no causally relevant factor has been omitted* from the initial set of possible factors.

I will illustrate this "causal spread" with two examples, which will be used to motivate the approach that follows. The first is the causation involved in a man breaking a leg and the second the inference involved in interpreting an utterance.

Example 1. A man is distracted and falls off a small ledge onto a pavement. When he lands his leg breaks. What caused his leg to break? It could be attributed to many things: the hardness of the pavement; the weakness of his femur; the way he landed on the leg; gravity; the mass of his body; him falling off the ledge; the ledge itself; the height of the ledge; the distraction; or even the man's distractability. There seems to be no end to the number of factors one *could* include as causes of the fracture. Whether one *does* count each of these as causes is arbitrary from an absolute external viewpoint. It can depend on the extent to which we judge each of them as *unusual.* For example, if the ledge was there due to a freak subsidence we might say that this subsidence was the cause – if the ledge was normal (the side of some steps) but the distraction was exceptional (there was a couple making love in the middle of the street) we would say the distraction was the cause.

Example 2. Two people, Joan and Jill, are talking: Joan says "We'll go and have a friendly chat in a bar."; Jill replies "Yeah, right!" which is (correctly) taken to mean by Joan that Jill thinks that this is a *bad* idea and does *not* want to go. In what way was the negative message conveyed? In other words, what allowed Jill to infer the meaning of Joan's utterance? There could be many such factors: the tone of Jill's voice; that the peer group to which Jill and Joan belong always say "Yeah, right!" when they disagree; that Jill is pointing a gun at Joan; that they are both are locked away in jail and so the suggestion was impossible to carry out; that Jill had been neurotically repeating "Yeah, right!" over and over for the past two years since her sister died etc. The answer could have been any one of these or any combination of them. Even if many of these factors *were* present Joan may have only used one or two of them in her inference, the rest being redundant.

Our models of the world (physical or social) are distinctly limited constructs. We could not possible learn useful models of our world if we had to include all the *possible* causes. In practice, we have to restrict ourselves to but a few causes that we *judge* to be the significant. The means by which we reach such judgements can vary greatly depending on the circumstances (including our knowledge etc.).

In general (as developing human beings) we start by learning simple models of our world, i.e. those with only a few explicated causes and only introduce more causes as we need to. The more causes we include in our model the more generally applicable, but also the more unwieldy, it becomes. If we are lucky, the natural world is so structured as to allow us to abstract away some of this detail and find a more generally applicable model for certain aspects that are relevant to us. Sometimes we can construct models that have sufficiently wide conditions of application that it is convenient for us to *consider* them as general truths. However, such cases, are exceptional – they tend to be highly abstract and so to apply them one typically has to bring the cluttering detail back in to the model in the process of applying it to a particular situation. In many models in the field of physics, this detail is frequently bought in as either initial conditions or auxiliary hypotheses.

In this paper I want to consider aspects of the more usual models we learn and apply, not the exceptional ones that are we consider as generally applicable. There is a view that somehow more general models are *better*, because they are not restricted to particular domains of applicability, and hence should be the focus of our study. According to this view more specialised knowledge *should* be represented as specialisations of these 'general' models. I dispute this – I contend that although there is great theoretical economy of representation in the more abstract and generally applicable models, the huge difficulties of applying them to common situations often precludes them as a sensible way to proceed. It would be incredible indeed if it just so happened that the world was constructed so that it was *always* sensible to work via the most general structures possible!

3. Contexts emerge from Modelling Heuristics

The efficacy of our limited learning and inference in dealing with our complex world is dependent on the presumption that many of the possible causes of events that are important to us remain relatively constant. Otherwise we would need to include all the possible causes in our models and decision making processes, which would not be feasible. This relative constancy is what makes our limited modelling ability useful: we can learn a simple model in one circumstance and successfully use it in another circumstance that is sufficiently similar to the first.

Roughly, I am going to attribute the label of 'context' as a stand-in for those factors that are *not* explicitly included in the simple models we learn, or, to put it positively, those factors that we use to recognise when a model is applicable. This is similar to Zadrozky's approach:

> *"...for any procedure we can divide its parameters into two sets: those that change with each invocation of the procedure and those that are there but remain constant. The latter set will be called its content and the former its focus." [21]*

It is the possibility of the transference of knowledge via fairly simple models from the circumstances where they are learnt to the circumstances where they are applied which allows the emergence of context. The utility of 'context' comes from the possibility of such transference. If this were not feasible then 'context', as such, would not arise. This process of transference is illustrated below in figure 1.

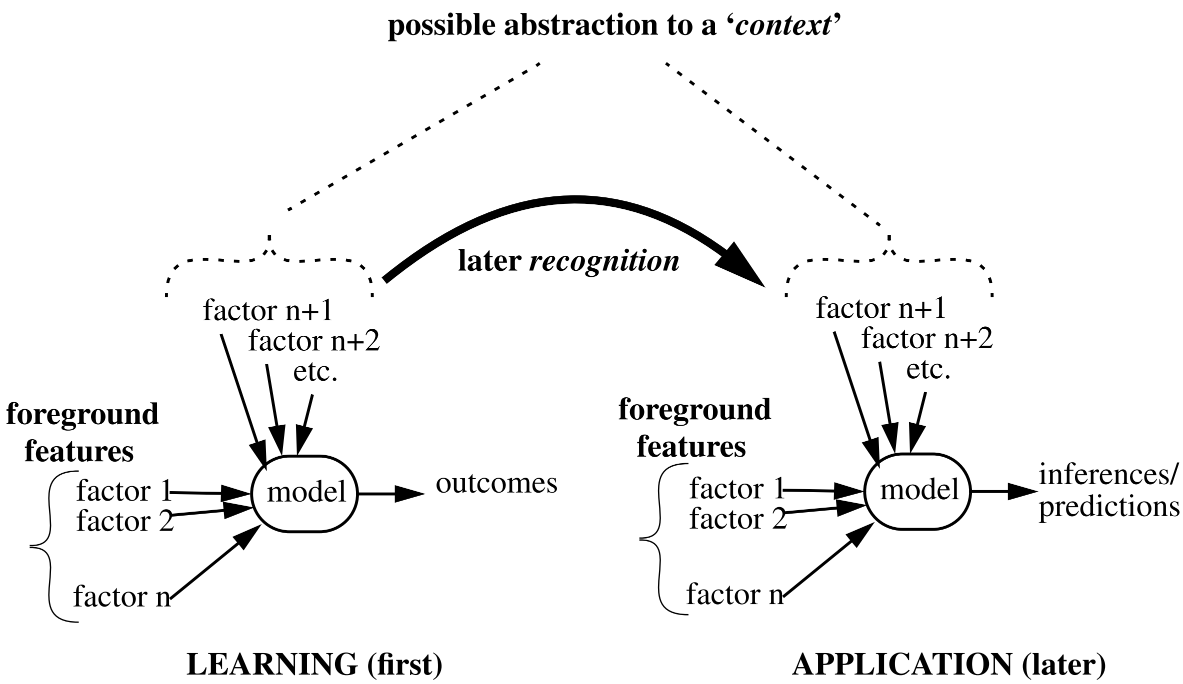

Figure 1. The use of context in the transference of knowledge between learning and application

For such a transference to be possible a number of conditions need to be met, namely:

- that some of the possible factors influencing an outcome are separable in a practical way;
- that a useful distinction can be made between those factors that can be categorised as foreground features (including 'causes') and the others;
- that the background factors are capable of being *recognised* later on;
- that the world is regular enough for such models to be at all learnable;
- that the world is regular enough for such learnt models to be at all useful when applied in situations where the context can be recognised.

While this transference of learnt models to applicable situations is the basic process, observers and analysts of this process might identify some of these combinations of features that allow recognition and abstract them as *a* 'context'. This usually is possible because the transference of knowledge as models requires that the agent doing the transference can recognise these characteristic combinations, so it is possible that an observer might also be able to do so and give these combinations names. Note that it is not *necessarily* possible that such an observer will be able to do this as the underlying recognition mechanism may be obscure. Of course, it may be that the agent doing the transference *itself* analyses and abstracts these features, and thus makes this abstract available for reflective thought.

Given the above conditions are possible, I am defining 'context' as:

> *the abstraction of those elements of the circumstances in which a model is learnt, that are not used explicitly in the production of an inference or prediction when the model is*

later applied, that allow the recognition of new circumstances where the model can be usefully applied.

Due to the fact that context is characterised as an abstraction of an aspect of a heuristic for the learning *and* application of knowledge, the properties of such contexts can not be meaningfully analysed if one only considers *either* the learning *or* the application of such knowledge. If one did this one would not only be missing out on over half of the story but also undercutting the reasons for its very existence. If the problems of learning are ignored then there is no reason not to encode such models without context – the non-causal factors can be treated as either given or the same as the other features of the model, de-contextualising them. If the problems of inference are ignored then there is no reason to separate the recognition of an appropriate context from that of recognising the correct prediction in that context.

4. Internal and External Conceptions of 'context'

Given the above picture of context and ignoring, for the moment, the effect that different heuristics will have in different domains, there are at least two ways in which we (as people discussing the idea of context) can make a reference to this process. I will call these the 'internal' and 'external' ways of referring to context. The distinction drawn here is not new (e.g. [10]), but I wish to re-tell it in terms of the picture presented herein.

4.1. The internal approach

We can refer to the context as that which an individual (or group of individuals) *actually* uses as a result of their learning. This has the disadvantage that different individuals or groups will develop different constructs as a result of their circumstances and the heuristics they happen to use. On the other hand these can be empirically investigated.

It is not clear that the contextual mechanism that an individual uses to remember and recognise a situation will be best represented by symbolic inference. For it may be that one such 'context' isn't *clearly* separable from another. Deciding which context is relevant may be more of a process of recognition than an inferential process. If this is the case the recognition might be better modelled by something like a neural network than using a logic-based system. It may be that there isn't sufficient continuity for the results of the recognition process to be meaningfully ascribed separate identities. But even if this *is* the case, it does not mean that it is useful for *us* to analyse and model these mechanisms using computational, symbolic or other reified terms.

4.2. The external approach

Alternatively, we could try to abstract from the concrete manifestation of individual's constructs outwards to the regularities and features these constructs rely on in order for their modelling heuristics to be useful (or even possible). Thus we have talk of context-as-a-resource in NL [11], or the context-we-inhabit in AI [3].

The problem with this approach is that the number of possible outward features that *might* be useful can be large. In order to focus on the parts that might actually be useful for an intelligence that is attempting to exploit them one has to consider (maybe implicitly) the internal construct of a context anyway. There may be some good grounds for identifying some relevant regularities on a priori grounds (for example, temporal context) but even in these cases it is hard to see how the properties of such contexts could be deduced for *actual* agents in real examples, without some validation that the presumed a priori grounds were *actually* used by the agent.

Thus in each case the pragmatics of learning, transferring and applying knowledge creeps in. The only escape from the relevance of these pragmatic roots of context is if one is not considering an *actual* or *applicable* rationality and reasoning but only some artificial rationality for use on problems in restricted domains (e.g. a heavily idealised or purely normative rationality).

5. Context in Different Domains

Different modelling heuristics will be useful in different domains, which explains why different sorts of context arise in these different domains. The modelling heuristics typically exploit some sort of commonality. This commonality ensures that some of an event's features will remain constant from the time of learning a model to its application. This commonality makes the modelling of events feasible by limiting the number of features have to be explicitly included in the model, under the conditions that the commonality is either pervasive or recognizable.

Sometimes these common features can be identified, and the external approach to context adopted (as discussed in section 4.2 above), but at other times this may not be obvious so that one is forced to indentify the heuristics that happen to be used by the agent, leading to the internal approach (section 4.1). I list three broad areas of commonality below, and discuss the likely tractability of the contexts that may arise from them as an objects of study.

5.1. Shared physical environment

One of the most obvious and straight-forward commonalities is a shared spatial location or time. This can either mean that a model is learned at a particular location and time and then applied at similar location or time (e.g. on Sunday in church), or it can mean that there is a spatial and temporal commonality between conversers allowing a listener to infer the meaning of indexicals. Such physical commonalities are, by their nature, readily indentifiable. This means that laws of spatial and temporal context are among the most amenable of all contexts to analyse. For an introduction to the situation calculus approach for this type of commonality see [2].

5.2. Shared social environment

The richest source of commonality we have as humans, arises from our shared culture, in particular our language. In fact the trouble is that often, this source is *too* rich for us as academics. As Graeme Hirst points out there are often *no* external features that are identifiable as the commonality, since the commonality may be a purely social construct with no accompanying external markers [11]. This would mean that the external approach would not be viable. Of course, in such situations the internal approach is almost as difficult, since the heuristics used by individuals may vary amoung individuals at the same time and over time for one individual the same. This embarrassment of riches may well mean that there are no *general* characteristics that can be abstracted from the multitude of heuristics used to model these social constructs, and hence no generally applicable characteristics of social context *per se*.

This does not mean that the relevant commonalities and heuristics can not be discovered in *particular instances*. For example, if a set of social norms has been established within a certain social group, then it might be sensible (in circumstances where one recognises that the situation lies within that grouping) to model others' behaviour in terms of deviations from these norms. If these norms have been sufficiently externalised into an explicitly expressed set of rules this will be identifiable as a source of social context. Akman outlines some other feasible approaches to aspects of social context in [1].

5.3. Shared biology

A third area of commonality is our shared biology. This may provide a shared experience of emotions, experience of inhabiting a body, experience of consciousness and other shared knowledge (e.g. basic language structure). These may be very important in a child's early development but are complicated and 'masked' with cultural overlays in later life. However their relative constancy across humans and their pervasiveness may allow for studies of context arising from these commonalities in a way that is difficult with commonality based in social constructions. Apart from linguists who apply the concept of deep-structure, I do not know of any studies of context which focus in these biological areas[2].

6. Bottom-up and Top-down Approaches to Modelling Context

In addition to the differences in context resulting from the heuristics relevant to a domain and whether an 'internal' or 'external' approach is taken, there are also differences that are imposed by us as modellers due to the different approaches we use for investigating these situations. One of the most important (in terms of contrasts in approach) is whether one attempts to formulate one's models using a 'top-down' or 'bottom-up' approach. A top-down approach is where one attempts to lay down general principles (encoded variously as axioms, rules, algorithms, etc.) based on current or *a priori* thought. A bottom-up approach would involve attempting to induce models of context from the details of the learning and inferential processes as they might occur in practice, and later seeing if any appropriate abstractions or generalisations suggest themselves.

6.1. Top-down

The approach to modelling context most frequently taken in AI, and perhaps epitomised by the approach of John McCarthy [12], which is to specify a general structure for representing statements concerning contextual reasoning and then to investigate some of the possible axiomatisations of logics that encapsulate principles that are thought desirable. Thus the general principles are formulated first and the properties emerging from these are investigated later. The initial standard for judging such constructs is the plausibility and generality of the abstract principles – thus like mathematics this is a foundationalist approach. Of course, the *ultimate* judgement comes from the usefulness of the approach in formalising or implementing actual systems. This approach is partly a result of a desire to elucidate generally applicable AI principles and partly a bias resulting from the selection of formal logics as a tool for modelling practical reasoning.

Of course, some work in AI takes a less general approach than this, especially where the work is focused towards a specific problem or problem domain.

6.2. Bottom-up

In other domains it is much more difficult to establish general principles for learning and inference. Here a more bottom-up approach needs to be taken. A small sub-domain is typically chosen and then relevant examples considered to establish the likely heuristics involved. In this approach the specific data

2. Although Drescher's 'schema' touch on this area [5].

and facts come first and the more abstract principles and theories come second. The models are formulated to capture or explain observed processes and will be judged in this light. Later more abstract models (or laws) might be posited from testing against these models and the data. Thus this approach could be dubbed the ‘scientific’ approach.

The bottom-up approach is perhaps taken most seriously by those who advocate a constructivist approach to AI. Here care is taken to assume *as little as possible* in advance so that as much as possible of the behaviour is available for capture in the models induced [17].

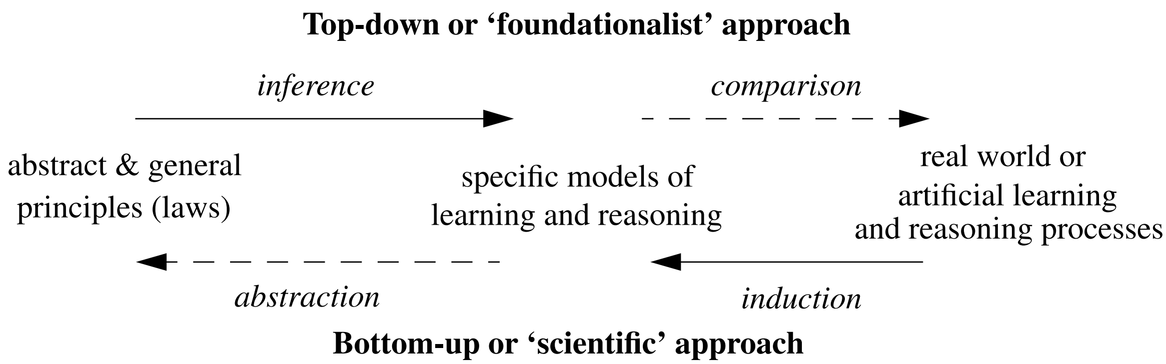

Figure 2. Top-down and bottom-up approaches to the investigation of context

The community interested in context is unevenly split into these two approaches. The ‘foundationalists’ are searching for a sort of *mathematics* of context, their approach is principle-based and has a potentially general applicability. However they are dogged by problems of scalability from the toy-problems they are tested on and will only be as strong as their *a priori* principles turn out to be. The ‘scientists’ are typically working in a specific domain with more realistic processes and problems, they face the same difficulties of generalisation as other scientists – it is a slow and difficult process to discover successful theories. On the other hand any progress they make will be strongly grounded in real processes and have clear conditions of applicability. In other sciences both kinds of approach have turned out to be useful, the foundationalists have typically had a role in producing a palate of formalisms, a few of which turn out to be useful to the scientists who use them for describing or modelling the actual phenomena[3]. The high odds against a particular type of formalism turning out to be useful means that it is vital that the *maximum possible* variety of approaches be developed. The scientists have the job of finding the mappings from the phenomena concerned to models expressed in these formalisms. This job is harder, which possibly explains why they are in the minority.

7. A Bottom-up Investigation of Internal Context in an Artificial Agent

To illustrate how a bottom-up investigation of context might proceed, I exhibit a simulation which allows the analysis of the emergence of context in the knowledge learned by an artificial agent in a controlled environment. The model and results are preliminary and are intended to be more suggestive of future methodology than significant in detail.

3. They have also had a less glorious, but still important, role of setting up the straw men for the ‘scientists’ to knock down!

7.1. Overview

The idea is to place an artificial agent in a environment that is well beyond its capacity to model but one which exhibits some regularities. The agent is designed to learn about its environment using feedback from its attempts to predict outcomes in that environment. It learns using a method that allows the development of pattern-recognition, inference and context-like constructs. I then study the structure of the knowledge that the agent learns in order to identify whether context-like constructs emerged.

7.2. The Learning Algorithm

The learning algorithm is based upon the neural network invented by Chialvo and Bak in [4]. This algorithm learns a set of mappings between single inputs and single outputs using purely negative feedback. The information is stored as the set of weights associated with the arcs connecting a set of nodes. The algorithm works as follows: an input node is fired, then the arc from this node with the greatest weight fires which fires the node it leads to; this carries on until an output node is fired; if the output is correct then nothing happens but if it is wrong all the weights on fired arcs are depressed and this change in weight is redistributed to other arcs.

This ‘learning by mistakes’ algorithm is very efficient to train and use. Also, due to the fact that successful associations are not positively reinforced, the network remains in a critical state so that if the environment it is learning about returns to a previous state, the associations it had learned are relatively quickly re-established. In other words, since an arc either fires or it does not, the network only needs to marginally depress previously correct associations and hence does not totally ‘forget’ them. The topology of the network turns out not to be critical for the working of this algorithm. This algorithm is illustrated in section 3.

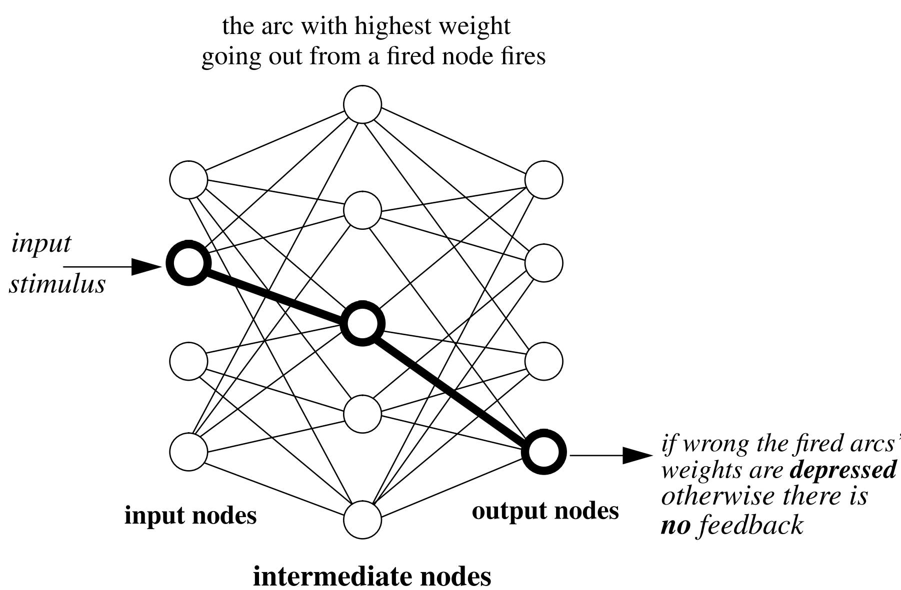

Figure 3. Chialvo and Bak’s learning algorithm

I have adapted this algorithm in two ways: *firstly*, to allow the network to learn mappings from combinations of inputs to combinations of outputs and*, secondly*, to add ‘switching’ arcs that can turn other arcs on or off.

The first adaption was achieved by introducing a global critical level for the network. Arcs leading from a fired node can only fire if their weight is greater than the critical level. The critical level is gradually changed so as to ensure that the appropriate level of firing occurs. As before, all fired arcs that cause output nodes to fire when they should not have are depressed.

Secondly, arcs are of two kinds: switched and unswitched. Unswitched are as already described, switched arcs are enabled by an arc that leads to them (an *enabling* arc)– they can only fire if: (1) the nodes they come from have fired; (2) their weight is greater than the critical level; *and* (3) the arc that leads to them has fired. If the firing of an enabling arc leads to the over-firing of an output node then it is depressed in a similar manner to other arcs but to a lesser degree (e.g. depressed by 25% of the value by which the other arcs are depressed). This algorithm is illustrated in figure 4.

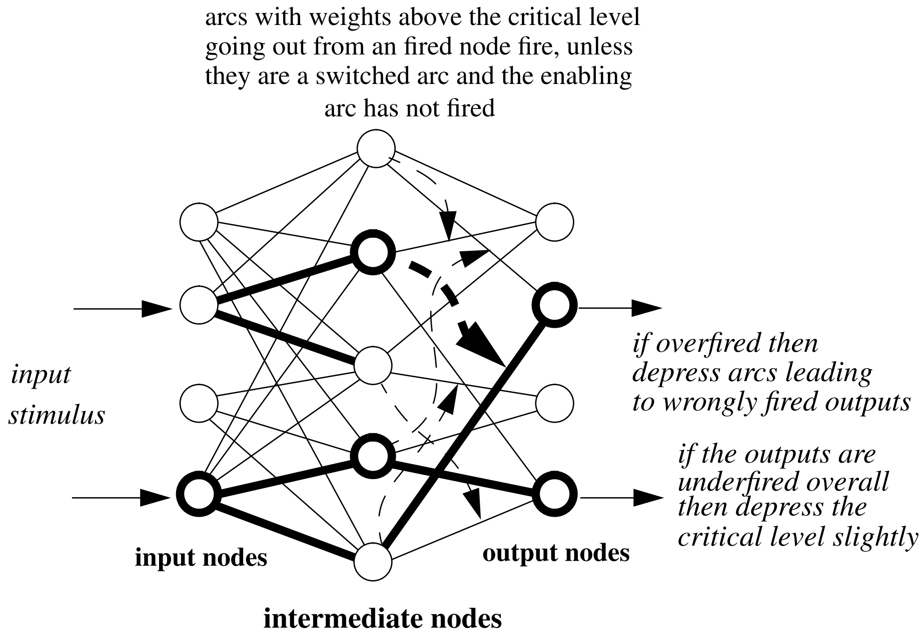

Figure 4. Generalised version of the algorithm

7.3. Analysis of the Network in Terms of Learned Knowledge

This generalised version of Caviallo and Bak's algorithm is *not* designed to be a particularly effective learning mechanism but rather a tool for investigating what sort of rules are learnable in an environment. It is particularly appropriate for this task for two reasons: *firstly*, because the network can be readily analysed since the arcs that have weights greater than the critical level can be interpreted as implications; and *secondly*, because the switched arcs allow the emergence of nodes that act as contexts, in that they do not directly cause the firing of further nodes but *enable* a set of other implications without this being imposed.

The network is designed so that it can learn the structures of directed arcs described in [9]. It is designed to be as free from assumptions about the structure of the contexts as possible – thus it is ideal for this kind of investigation where the purpose is to investigate *what the appropriate assumptions are* in a particular environment. Broadly speaking, a context is represented by one (or more) nodes that develop a role of 'switching' sets of associations, whilst other nodes represent facts about the environment. There is, of course, no hard and fast distinction between context nodes and other nodes but more that in some circumstances some nodes act more *as* contexts and others act more *as* the content of the model in context. The difference is illustrated in figure 5.

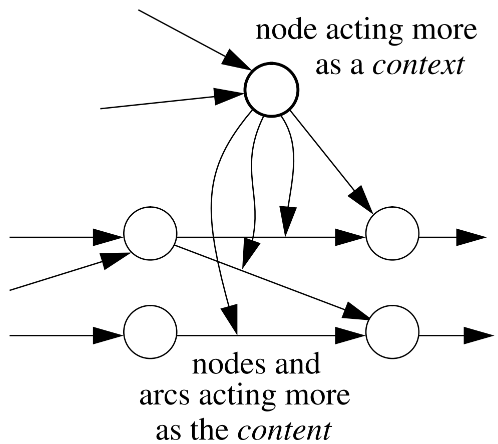

Figure 5. Nodes acting as contexts and other nodes

If the structure of the network allows it, there is nothing in the algorithm that prevents the network: learning in a context-free way; having contexts imply other contexts, developing hierarchies of contexts, having nodes acting as contexts in some situations and not in others etc. It can be interpreted as implementing both inferential and pattern recognition processes: whether a node is fired is a matter of pattern recognition as a result of the learning done by the network, but the resulting firable arcs can be analysed in terms of (possibly context-dependent) implications about its environment.

7.4. The Environment

The environment is a small artificial stock market. Broadly it is an extension of [14]. There is a small population of artificial traders who are each attempting to predict the next day's stock prices based on data from the recent past. Each trader has a small population of candidate models of the price behaviour in terms of past movement, dividends, comparisons against historical levels etc. Each agent is continually: generating new variations of past models; evaluating them as to how well they would have predicted prices in the recent past; choosing the best of its models; using this model to predict future prices and taking an action to buy or sell dependent on this prediction. It is doing this in parallel to the other agents – essentially each is trying to 'out-think' the other traders by predicting the effect of each other's predictions. The price series that results are weakly related to the fundamentals of the stocks (e.g. the dividend to price ratio) but this is 'masked' by self-reinforcing patterns of speculation among the traders.

I enhanced this model by substituting a genetic programming algorithm for the classifier system. This enables the traders to develop their internal models in a more open-ended and creative way using a much wider range of strategies, including imitation. I also introduced a round of communication between traders before each session of buying and selling. The details of the traders and their interactions can be found in [6, 7], but they are not critical here. What is important here is that it provides: an environment that is beyond the capacity of agent to completely model [8]; that displays distinct phases to which learning heuristics might apply; and that is tunable in the level and type of learning difficulties it presents. To give an idea of the level of difficulty, figure 6 shows a typical example of the price series that the agent is trying to learn. Although it is fairly unpredictable *in detail*, there are distinct and recognizable phases of buying and selling indicated by the cyclical nature of price swings and the clear negative correlation between the prices of the two stocks.

7.5. Implementation

The whole model was implemented in SDML, a declarative modelling language developed at the Centre for Policy Modelling for social simulation. For details of SDML see [13] and http://www.cpm.mmu.ac.uk/sdml.

7.6. Preliminary Results

Since the purpose of this model is merely to demonstrate the feasibility of a bottom-up approach, this description of the results will be cursory. Those who want more detail will either have to imitate the techniques or wait for a fuller investigation of the subject by me.

Two runs were performed with the same learning agent but different environmental set-ups. The first run's environment consisted of a market with 5 traders, each of which had 20 models of initial depth 5, while the second had only 2 agents each with 10 models of initial depth 4. The first will be called the 'harder' and the second the 'easier' learning task. The prices series that results from these are illustrated in figure 6 and figure 7 respectively. The 'harder' series change more abruptly and exhibit a more complex cycle than the 'easier' series.

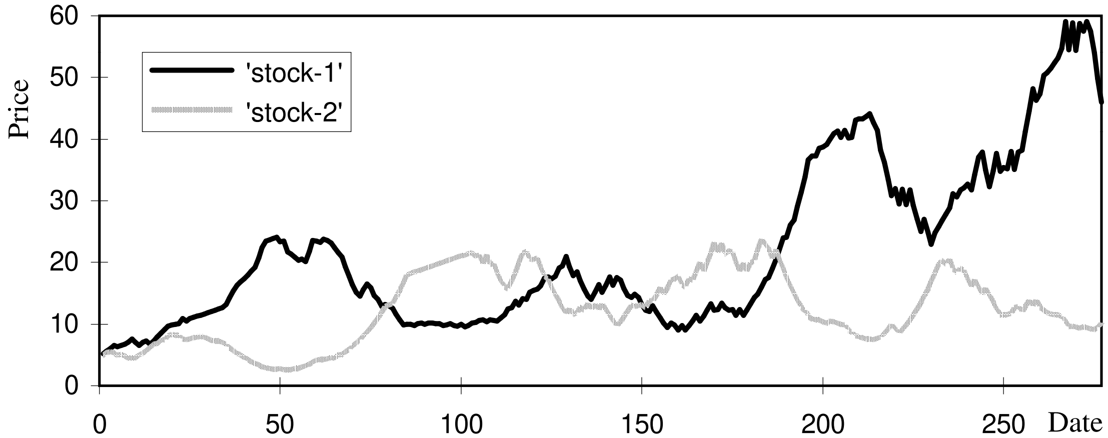

Figure 6. A price series output from the market (two stocks) – harder learning task

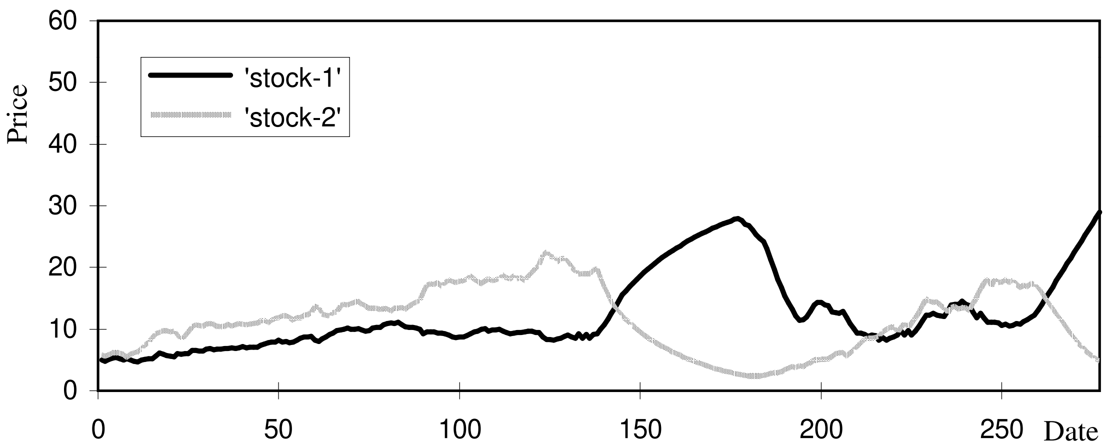

Figure 7. A price series output from the market (two stocks) – easier learning task

For each task the agent was given an identical network structure of: 23 input nodes, 20 intermediate nodes; and 11 output nodes. The input nodes were randomly connected to an average of four intermediate nodes each, which were connected with 3 output nodes and switching 3 other arcs using enabling arcs as described above. The agent then runs through the dates once, hence it is a one-shot incremental learning task. The network is deliberately limited in order to simulate the sort of learning a real agent might attempt in a similar situation.

Despite the considerable difficulties it was faced with the agent did manage to induce some effective (but, of course, fallible) rules about the prices series. For both tasks the agent specialised some of the nodes to act as indications of context, that is it learnt arc weights such that these nodes acted to 'switch' other arcs and did not imply any of the output nodes directly themselves. In figure 8 I have plotted the number of these 'context' nodes as the simulation progressed for both learning tasks. It can be seen that the agent induced more contexts for the 'easier' task than the 'harder' one.

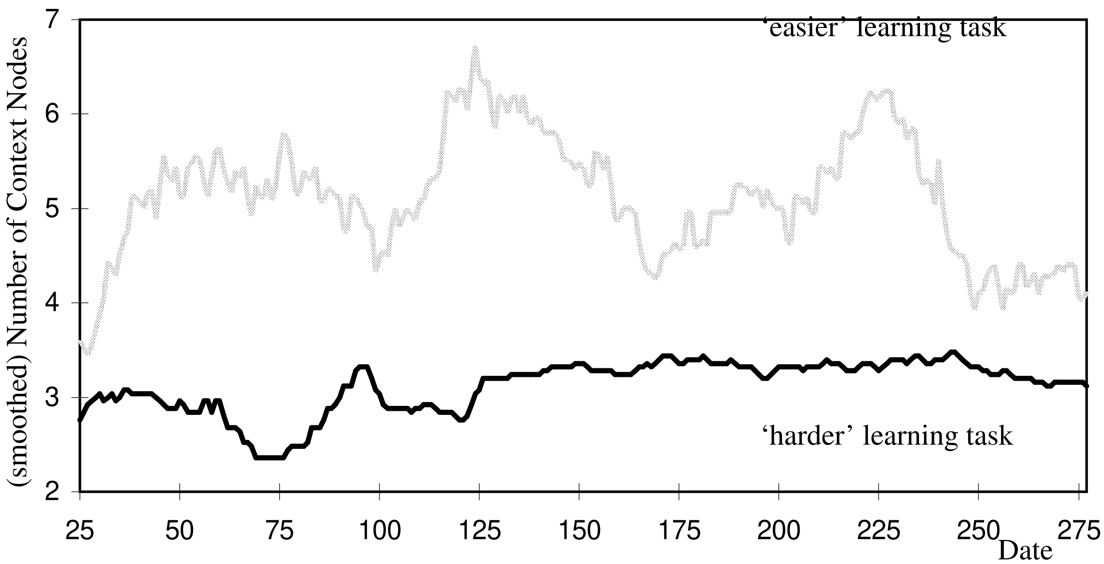

Figure 8. Number of nodes acting as contexts – both learning tasks

In figure 9 I have summarised the contexts developed by the agent at the end of the 'harder' learning task. These were constantly being selected in and out, due to the nature of the learning algorithm. The underlined headings indicate the conditions under which the context is recognised and the implications under them are the rules that are hold in that context (i.e. are enabled by them). Thus the consequent of one of these rules will only be inferred if the context is fired as well as the antecedent, but this does not mean that the context is functionally the same as the antecedent as they are reinforced by the learning algorithm in different ways. The enabling arcs' weights are changed at a slower rate than the node-node arcs (one quarter as slowly). Thus it will take many more occasions of negative feedback to change a context node's associations than one of the implicational arcs in that context. This allows learnt contexts to last longer. Likewise the arcs that lead to the contextual node (i.e. those representing the conditions which the context is recognised) can change more quickly than the 'membership' of that context.

In this way I have allowed the emergence of context-like behaviour without imposing the sort of two-tier structures that have been employed in other machine learning algorithms (e.g. [16, 20]). If I had a network with more than one intermediate layer we could allow the emergence of contexts within contexts etc. but this would take longer runs in order for any results to emerge.

last(Stock-2 unchanged)
last(Index unchanged) → Stock-2 down
last last(Stock-2 up) → Stock-2 unchanged OR Stock-2 down

last(Stock-2 up)
last(Index unchanged) → Stock-2 unchanged

True
last(Index down) → Stock-2 down

last last last(Index down)
last(Stock-2 down) OR last last(Index down) OR last last(Stock-1 unchanged) → Stock-1 up

last(Index high)
last last(Index unchanged) OR last last(Index up) OR last(Index down) → Stock-2 unchanged
Index up

last last(Index down) OR last(Stock-1 unchanged) OR last(Stock-2 down)
last last(Index down) OR last(Index high) → Stock-1 up

last last(Index up) OR last last(Index unchanged)
last(Index high) → Stock-2 unchanged
Index unchanged

Figure 9. The 'contexts' induced by the agent by the last 10 dates – harder learning example

Future work with this model will involve attempting more runs of the current model; runs with a more intricate network topology and runs on data generated by real-world processes. The idea is not to discover any 'golden rules' for these domains but to start to tease out how and in what way different contextual structures can arise from different sorts of regularity (or lack of it) in the data.

8. Conclusion

If one only studies *either* the learning of context-dependent knowledge *or* context-dependence inference then one may well be missing the essence of context. I suggest that context only makes complete sense when considering the *transference* of knowledge from point of learning to point of application. Identifiable contexts arise from *our modelling* of those features that allow the recognition of a situation in which an inferential model can be applied. If this is correct then a successful study of context may need the *combined* expertise of the AI and Machine Learning communities.

Acknowledgements

Thanks to Scott Moss, Varol Akman and Helen Gaylard for comments on these ideas, to Pat Hayes for stimulating arguments about context and related matters and Steve Wallis for writing SDML.

SDML has been developed in VisualWorks 2.5.1, the Smalltalk-80 environment produced by ObjectShare. Free distribution of SDML for use in academic research is made possible by the sponsorship of ObjectShare (UK) Ltd. The research reported here was funded by the Economic and Social Research

Council of the United Kingdom under contract number R000236179 and by the Faculty of Management and Business, Manchester Metropolitan University.

References

[1] Akman, V. (1997). Context as a Social Construct. Context in Knowledge Representation and Natural Language, AAAI Fall Symposium, November 1997, MIT, Cambridge.

[2] Akman, V. and Surav, M. (1996). Steps Towards Formalizing Context. *AI Magazine*, 17:55-72.

[3] Barwise, J. and Perry, J. (1983). Situations and Attitudes. Cambridge: MIT Press.

[4] Chialvo, D. R. and Bak, P. (1997). Learning by Mistakes. Sante Fe Working Paper 97-08-077.

[5] Drescher, G. L. (1991). *Made-up Minds – A Constructivist Approach to Artificial Intelligence.* Cambridge, MA: MIT Press.

[6] Edmonds, B. (1998). Modelling Socially Intelligent Agents. *Applied Artificial Intelligence*, 12.

[7] Edmonds, B. (forthcoming). Modelling Bounded Rationality In Agent-Based Simulations using the Evolution of Mental Models.In Brenner, T. (ed.), *Computational Techniques for Modelling Learning in Economics*, Kluwer.

[8] Edmonds, B. (forthcoming). Capturing Social Embeddedness: a Constructivist Approach. *Adaptive Behaviour.*

[9] Edmonds, B. A Simple-Minded Network Model with Context-like Objects. European Conference on Cognitive Science (ECCS'97), Manchester, April 1997. (http://www.cpm.mmu.ac.uk/cpmrep15.html)

[10] Hayes, P. (1995). Contexts in Context. Context in Knowledge Representation and Natural Language, AAAI Fall Symposium, November 1997, MIT, Cambridge.

[11] Hirst, G. (1997). Context as a Spurious Concept. Context in Knowledge Representation and Natural Language, AAAI Fall Symposium, November 1997, MIT, Cambridge.

[12] McCarthy, J. (1996). A logical AI approach to context. Unpublished note, 6 February 1996. http://www-formal.stanford.edu/jmc/logical.html

[13] Moss, S., Gaylard, H., Wallis, S. and Edmonds, B. (1998). SDML: A Multi-Agent Language for Organizational Modelling. *Computational and Mathematical Organization Theory*, 4, 43-69.

[14] Palmer, R.G. et. al. (1994). Artificial Economic Life – A Simple Model of a Stockmarket. *Physica D*, 75:264-274.

[15] Pearl, J. (forthcoming). An Axiomatic Characteriztion of Causal Counterfactuals. *Foundations of Science*.

[16] Trun S. and Mitchell, T. M. (1995). Learning One More Thing. Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95). San Mateo, CA: Morgan Kaufmann, 1217-1223.

[17] Vaario, J. (1994). Artificial Life as Constructivist AI. *Japanese Society of Instrument and Control Engineers, 33:65-71*.

[18] Wagner, A. (1997). Causality in Complex Systems. Sante Fe Working Paper 97-08-075.

[19] Wheeler, M. and Clark, A. (forthcoming). Genic Representation: Reconciling Content and Causal Complexity. *British journal for the Philosophy of Science*.

[20] Widmer, G. (1997). Tracking Context Changes through Meta-Learning. *Machine Learning*, 27:259-286.

[21] Zadrozny, W. (1997). A Pragmatic Approach to Context. Context in Knowledge Representation and Natural Language, AAAI Fall Symposium, November 1997, MIT, Cambridge.