Capturing Social Embeddedness: a constructivist approach

Bruce Edmonds
Centre for Policy Modelling,
Manchester Metropolitan University

Abstract

A constructivist approach is applied to characterising social embeddedness and to the design of a simulation of social agents which displays the social embedding of agents. Social embeddedness is defined as the extent to which modelling the behaviour of an agent requires the inclusion of the society of agents as a whole. Possible effects of social embedding and ways to check for it are discussed briefly. A model of co-developing agents is exhibited, which is an extension of Brian Arthur 's 'El Farol Bar' model, but extended to include learning based upon a GP algorithm and the introduction of communication. Some indicators of social embedding are analysed and some possible causes of social embedding are discussed.

Keywords: simulation, embedding, agents, social, constructivism, co-evolution

1 Introduction

In the last decade there has been a lot of attention paid to the way the physical situation of a robot affects its behaviour*. This paper focuses on the analogous importance of the *social* situation to an agent. It aims to identify phenomena that may be usefully taken to indicate the extent to which agents are socially embedded. In particular, it aims to do this for an artificial simulation involving co-evolving agents. In order to do this a modelling approach is adopted which takes ideas from the several varieties of constructivism.

The first section presents a brief overview of constructivism and its relevance to simulations of social agents. Then there is a section discussing the idea and possible effects of social embeddedness. A model illustrating differing degrees of social embeddedness is then exhibited. Both some general results and a couple of more detailed case studies are then presented. The paper ends with a short discussion of the possible causes of social embeddedness.

2 Constructivism and AI

Constructivism, broadly conceived, is the thesis that knowledge can not be a passive *reflection* of reality, but is more of an active *construction* by an agent. Although this view has its roots in the ideas of Kant, the term was first coined by Piaget [27] to denote the process whereby an individual constructs its view of the world. Extrapolating from this is Ernst von Glasersfeld's 'radical constructivism' [19] which approaches epistemology from the starting point that the *only* knowledge we can ever have is so constructed. In cybernetics it was used by Heinz Von Foerster [17], who pointed out that an organism can not distinguish between perceptions of the external world and internally generated signals (e.g. hallucinations) on *a*

priori grounds, but retains those constructs that help maintain the coherence of the organism over time (since those that do not will have a tendency to be selected out)*.

For the purpose of this paper the important aspects of constructivism are the following:

- the constructs are frequently not models, in the sense that they do not necessarily reflect the structure of agent's environment (as viewed by an external observer) – rather the constructs are merely compatible with the environment and the agent's existence in that environment;
- the constructs are closely related to the needs and goals of the agent, particularly in respect to its attempts to control its own actions and that of its environment;
- the constructs are built up as a result of frequent and active interaction with its environment rather than as a result of passive observation and reasoning;
- it emphasises the bottom-up approach to model building, with a tendency away from *a priori* considerations regarding cognition or rationality;

Constructivism has been taken up by some researchers in artificial intelligence and artificial life (e.g. [9, 28, 29]) as an approach to building and exploring artificially intelligent agents from the bottom up. Here, instead of specifying an architecture in detail from *a priori* considerations, the mechanisms and cognition of agents are developed using self-organisational and evolutionary mechanisms as far as possible. For this approach to be viable the agents must be closely situated in its target environment, since is it the serendipidous exploitation of features of its environment and the strong practical interaction *during* development which makes it effective (this distinguishes it from a lot of work in 'Artificial Life'). This is in contrast to what might be called an 'engineering approach' to artificial agents, where the agents are designed and set-up first and *then* let loose to interact with other such agents in order to achieve a specified goal. Constructivism in AI can be seen as an extension of the work of Rodney Brooks [5], but instead of the development of the organism happening through a design and test cycle done by human designers, the development is achieved via self-organisational and evolutionary processes acting on an agent situated in its environment.

This paper is constructivist in three different ways. *Firstly,* the approach to characterising social embeddedness is through properties of our constructs of the systems we are investigating. *Secondly,* the exhibited model is built in a constructivist AI style, in that: the content and development of an agent's cognition is specified as loosely as possible, where constructs are grounded in their effect upon the agent in conjunction with other agent's actions; and also that the meaning of the agent's communication is unspecified, so the effect of such communication is grounded in its use in practice and its development in the language-games that the agents appear to play. *Lastly,* constructivism is posited as a sensible explanation of the observed behaviour of the agents in the model described and hence, by analogy, as a possible explanatory tool for other social situations.

3 Social Embeddedness

3.1 Characterising Social Embeddedness

In attempting to elucidate the concept of ‘social embeddedness’, one faces the problem of where to base one’s discussion. In sociology it is almost an assumption that the relevant agents are ultimately embedded in their society – phenomena are described at the social level and their impact on individual behaviour is sometimes considered. This is epitomised by Durkheim, in that he claims that some social phenomena should be considered entirely separately from individual phenomena [10]. Cognitive science has the opposite perspective – the individual’s behaviour and processes are primitive and the social phenomena may emerge as a *result* of such individuals interacting.

This split is now mirrored in the world of computational agents. In traditional AI it is the individual agent’s mental processes and behaviour that are modelled and this has been extended to considerations of the outcomes when such autonomous agents interact. In Artificial Life and computational organisational theory the system as a whole is the focal point and the parts representing the agents tend to be relatively simple.

I wish to step back from disputes as to the extent to which people (or agents) *are* socially embedded to one of the appropriateness of different types of models of agents. From this view-point, I want to say that an agent is *socially embedded* in a collection of other agents to the extent that it is more *appropriate* to model that agent as part of the total system of agents and their interactions as opposed to modelling it as a single agent that is interacting with an essentially unitary environment. Thus I have characterised social embeddedness as a *construct* which depends on ones modelling goals, since these will affect the criteria for the appropriateness of models. It contrasts modelling agent interaction from an internal perspective (the thought processes, beliefs etc.) with modelling from external vantage (messages, actions, structures etc.). This is illustrated below in figure 1.

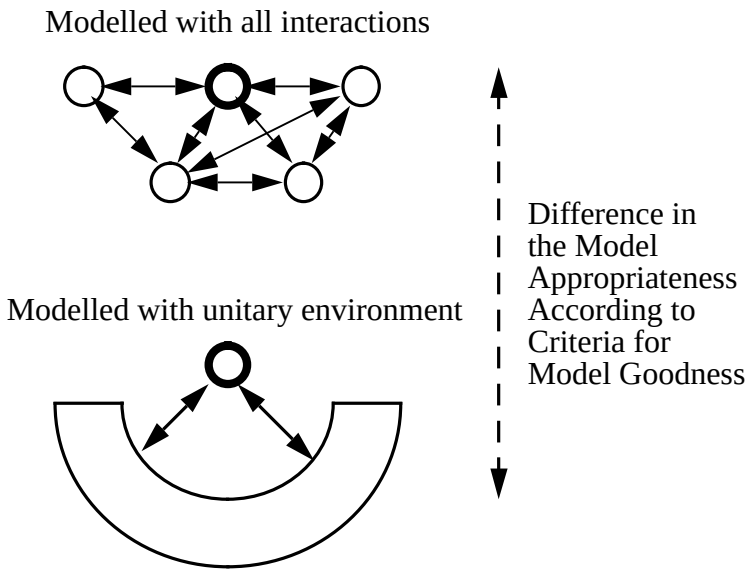

Figure 1. Social embeddedness as the appropriate level of modelling

This is not an extreme 'relativist' position since, if one agrees the modelling framework and criteria for model selection, the social embedding of agents within a collection of agents can be unambiguously assessed. Notice that criteria for model acceptability can include many things other than just its predictive accuracy, for example: *complexity* [12]. It is the *inevitability* of these other concerns which forces us to relativise this approach as one concerning the appropriateness of our constructs (along with the different modelling goals and frameworks). For example, a computer may be able to find obscure and meaningless models which (for computational purposes) separates out the behaviour of a single agent from its society (using something like genetic programming), which are totally inaccessible to a human intelligence. Also the modelling framework is indispensable; for example, an agent may not be at all embedded from an economic perspective but very embedded from the perspective of kinship relations.

Let us take a few examples to make this a little clearer.

- Consider first an economic model of interacting agents where each of these agents individually has a negligible effect on its environment, which would mean that a model of the whole system could be easily transformed into one of a single agent interacting with an economic environment*. Here one would say that each agent was not socially embedded since there is little need to model the system as a whole in order to successfully capture the agent's behaviour.
- Next, consider an intermediate case: an agent which interacts with a community via a negotiation process with just a few of the other agents. Here a model which just considers an agent, its beliefs and its interaction with these few other agents will usually provide a sufficient explanation for all that occurs but there may still be some situations in which interactions and causal flows within the whole community will become significant and result in surprising local outcomes. Here one could meaningfully attribute a low level of social embeddedness.

- Now consider the behaviour of an termite. It is possible to attempt to account for the behaviour of an termite in terms of a set of internal rules in response to its environment, but in order for the account to make any *sense* to us it must be placed in the context of the whole colony. No one termite repairs a hole in one of its tunnels only the *colony* of termites (via a process of stigmer gy: [20]). Here there is a significant level of social embedding.
- Finally, consider the phenomena of fashion. Something is fashionable only if considered so by a sizable section of a population. Although there is some grounding of the selection of particular styles in the climate, economic mood etc. this is tenuous. Fashion is lar gely a self-producing social phenomena; the *particular content* of fashion is contingent – it has little immediate connection with an individual' s needs, otherwise fashions would not vary so greatly or change so rapidly *. A model of how fashions change which took the form of an individual interacting with a *unitary* social environment would not capture much of the dynamics. Here we have a high level of embedding – fashion is an essentially social phenomena, so it is appropriate to model it at this level.

At first sight this seems a strange way to proceed; why not define social embeddedness as a property of the system, so that the appropriate modelling choices fall out as a natural result? The constructivist approach to characterising social embedding, outlined above, results from my modelling goals. I am using artificial agents to model real social agents (humans, animals, organisations etc.), and so it is not enough that the outcomes of the model are verified and the structure validated (as in [25]) because I also wish to characterise the emergent process in a *meaningful* way – for it is these *processes* that are of primary interest. This contrasts with the 'engineering approach' where the goal is different – there one is more interested in ensuring certain specified outcomes using inter-acting agents. When observing or modelling social interaction this meaning is grounded in the modelling language, modelling goals and criteria for model acceptability (this is especially so for artificial societies). The validation and verification of models can not be dispensed with, since they allow one to decide which are the candidate models, but most of the meaning comes from the modelling framework. The complexity of social phenomena (including, as we shall see in artificial societies) forces a 'pragmatic holism' upon us – that is, regardless of whether one is an *in principle* holist or an *in principle* reductionist, *in practice* we don't have the choice [11]. In simpler physical situations it may be possible to usefully attribute phenomena to an external reality but in social modelling we have to make too many choices in order to make progress. The proof of this particular pudding will ultimately be in the eating; whether this approach helps us obtain useful models of social agents or not.

The idea of social embedding is a special case of embedding in general – the 'social' bit comes from the fact we are dealing with collections of parts that are worthy of being called *agents*.

3.2 Possible Effects of Social Embeddedness on Behaviour

If one had a situation where the agents were highly embedded in their society, what noticeable effects might there be (both from a whole systems perspective and from the

viewpoint of an individual agent)? The efficacy of being socially embedded from the point of view of the embedded agent comes from the fact that if the most appropriate model is one that takes in far more than just its interactions with its social environment, then that agent will not have access to that model – it can not explicitly model the society it inhabits. In general, this may mean that:

- it will be more productive for the agent to cope by constructing behaviours that will allow it to exploit the environment rather than attempting to model its environment explicitly – in other words adopt an instrumentalist approach rather than a realist approach to its constructs, where the constructs are grounded in possible action*;
- as a result the constructs of an agent may appear somewhat arbitrary (to an external observer);
- it is worth frequently sampling and interactively testing its social environment to stand in stead of complete internal models of that environment (e.g. engage in gossip);
- agents specialise to inhabit a particular social niche, where some sub-set of the total behaviour is easier to model, predict, and hence exploit;
- at a higher level, there may be a development of social structures and institutions to 'filter out' some of the external complexity of its social environment and regularise the internal society with rules and structures (Luhman, as summarise in [3]);
- the agent's communications will tend to have their meaning grounded in their use in practice rather than as a reflection of an external social reality (since this inaccessible to the agent), thus their use of language might fit a Wittgensteinian analysis [31].

To summarise, the effect of being socially embedded might be that the agents are forced to construct their social knowledge rather than model that society explicitly.

3.3 Checking for Social Embeddedness

Given the presence of social embeddedness can have practical consequences on the modelled social behaviour, then it can be checked for. This is particularly so for a model of artificial agents, because the data is fully available. Given the approach to social embeddedness described above, it is necessary to specify the modelling framework and selection criteria first.

Let us suppose that our criteria for model goodness are complexity and explanatory power. By explanatory power, I mean the extent of the phenomena that the model describes. Thus there is a familiar trade-off between explanatory power and complexity in *our* modelling of our simulation [24]. If two descriptions of the simulation are functionally the same, the social embeddedness comes out as a difference between the complexity of the models at the agent and social levels†. This is not quite the obvious way of going about things – it might seem more natural to fix some criteria for explanatory power and then expand the complexity (in this case by including more aspects of the *social* nature of the environment in the model) until it suffices. However, in social simulation where it is often unclear what an acceptable standard of explanatory might be it is easier to proceed by making judgements as to the complexity of models.

In the model below we will use a rough measure of the social embeddedness based on where most of the computation takes place that determines an agent's communication and action. This will be indicated by the proportion of nodes which preform an external reference to the individual actions of other agents to those nodes that preform internal calculations (logical, arithmetic, statistical etc.). This ignores the computation due to the evaluation and production of the expressions inside each agent, but this is fairly constant across runs and agents.

4 A Model of Co-evolving Social Agents

4.1 The Set-up

The model is based upon Brian Arthur's 'El Farol Bar' model [2], but extended in several respects, principally by introducing learning and communication. There is a fixed population of agents (in this case 10). Each week each agent has to decide whether or not to go to El Farol's Bar on thursday night. Generally, it is advantageous for an agent to go unless it is too crowded, which it is if 67% or more of all the agents go (in this case 7 or more). This advantage is expressed as a utility, but this only impacts on the model in the agents evaluations of their constructs. Before making their decision agents have a chance to communicate with each other. This model can be seen as an extension of the work in [1], which investigates a three player game.

4.1.1 The environment

There are two alternative schemes for representing the utility gained by agents, which I have called: *friendly* and *crowd-avoiding*.

In the *crowd-avoiding* scheme each agent gets the most utility for going when less than 7 of the other agents go (0.7), they get a fixed utility (0.5) if they do not go and the lowest utility for going when it is crowded (0.4). In this way there is no fixed reward for any particular action because the utility gained from going depends on whether too many other agents also go. In this way there is no fixed goal for the agent's learning, but it is relative to the other agent's behaviour (which will, of course, change over time). Under this scheme it is in each agent's interest to discoordinate their action with the others (or, at least, a majority of the others).

The *friendly* scheme is similar to the *crowd-avoiding* scheme, there is a basic utility of 0.5 for going if it is not crowded, and 0.2 if it is but if they go to the bar each agents gets a bonus (0.2) for each 'friend' that also goes. If they stay at home they are guarenteed a utility of 0.65, so it is worth going if you go when it is not crowded with at least one other friend or if it is crowded with 3 or more friends. Who is a friend of whom is decided randomly at the beginning and remains fixed thereafter. Friendship is commutative, that is if A is a friend of B then B is a friend of A. An example of such a network is illustrated in figure 2. The number of friendships and nodes is constant accross runs but the detailed structure dif feres. In this scheme it is in the interest of agents to go when their other friends and only their friends are going. Under this scheme it is in each agent' s interest to coordinate its actions with its designated friends but to discoordinate its action with the other agents.

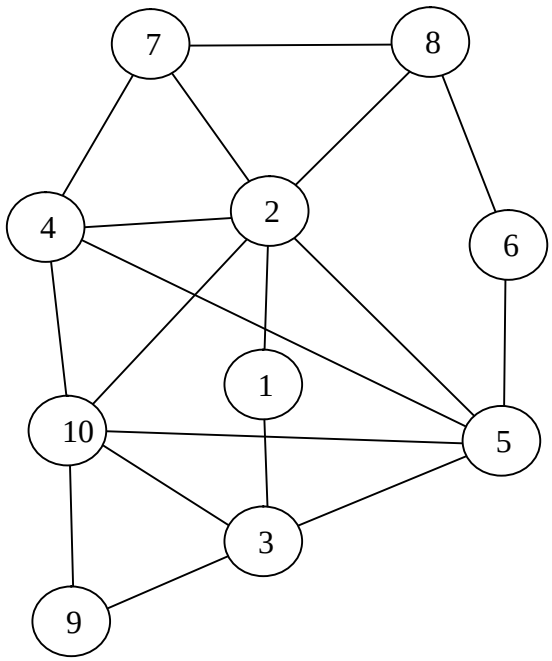

Figure 2. An imposed friendship network

Under both schemes it is impossible for all agents to gain the maximum utility, there is always some conflict to provide a potential for continual dynamics.

4.1.2 The agents

Each agent has a population of (pairs of) expressions that represent possible behaviours in terms of what to say and what to do (its constructs). This population is fixed in size but not in content. These expressions are taken from a strongly typed formal language which is specified by the programmer, but the expression can be of any structure and depth. Each agent does not 'know' the meaning or utility of any expression, communication or action – it can only evaluate each whole expression as to the utility each expression would have resulted in if it had used it in the past to determine whether it would go to the bar or not and the other's behaviours had remained the same. This is the only way in which the utilities affect the course of the model. Each week each agent takes the best such pair of expressions (in terms of its present evaluation against the recent past history) and uses them to determine its communication and action.

This means that any particular expression does not have an *a priori* meaning for that agent – any such meaning has to be learned. This is especially so for the expression determining the communication of the agents, which is only *implicitly* evaluated (and hence selected for) via the effect its communication has on others (and itself).

Each agent has a fairly small population of such models (in this case 40). This population of expressions is generated according to the specified language at random. In subsequent generations the population of expressions is developed by a genetic programming [21] algorithm with a lot of propagation and only a little cross-over.

The formal language that these expressions are examples of is quite expressive. The primitive nodes and terminals allowed are shown in figure 3. It includes: logical operators, arithmetic, stochastic elements, self-referential operations, listening operations, elements to copy the action of others, statistical summaries of past numbers attending, operations for looking back in time, comparisons and the quote operator.

Talk nodes:AND, OR, NOT, plus, minus, times, divide, boundedByPopulation, lessThan, greaterThan, saidByLast, wentLastWeek, randomIntegerUpTo, numWentLag, trendOverLast, averageOverLast, previous, quote

Talk terminals:IPredictedLastWeek, randomGuess, numWentLastTime

Action nodes:AND, OR, NOT, saidBy, wentLastWeek, previous

Action terminals:IPredictedLastWeek, IWentLastWeek, ISaidYesterday, randomDecision

Constants (either):1, 2, 3, 4, 5, 6, 7, 8, 9, 10, maxPopulation, True, False, barGoer-1, barGoer-2, barGoer-3, barGoer-4, barGoer-5, barGoer-6, barGoer-7, barGoer-8, barGoer-9 barGoer-10

Figure 3. The primitives allowed in the talk and action expressions

Some example expressions and their interpretations if evaluated are shown in figure 4. The primitives are typed (boolean, name or number) so that the algorithm is strictly strongly-typed genetic program following [23].

Talk expression:[greaterThan [randomIntegerUpTo [10]] [6]]

Action expression:[OR [ISaidYesterday] [saidBy 'barGoer-3']]

Interpretation: Say 'true' if a random number between 0 and 10 is greater than 6, and go if I said 'true' or barGoer-3 said 'true'.

Talk expression:[greaterThan [trendOverLast [4]] [averageOverLast [4]]]

Action expression:[NOT [OR [ISaidYesterday] [previous [ISaidYesterday]]]]

Interpretation: Say 'true' if the number predicted by the trend indicated by the attendance yesterday and four weeks ago is greater than the average attendance over the last four weeks, and go if I did not say 'true' yesterday or last week.

Talk expression:[OR [saidByLast 'barGoer-3] [quote [previous [randomGuess]]]]

Action expression:[AND [wentLastWeek 'barGoer-7'] [NOT [IwentLastWeek]]]

Interpretation: Say 'true if barGoer-3 said that last week, else say "[previous [randomGuess]]", and go if barGoer-7 went last week and I did not.

Figure 4. Some example expressions

The reasons for adopting this particular structure for agent cognition is basically that it implements a version of rationality that is credible and bounded but also open-ended and has mechanisms for the expression of complex social distinctions and interaction. In these respects it can be seen as a step towards implementing the 'model social agent' described in [6]. For the purposes of this paper the most important aspects are: that the agent constructs its expressions out of previous expressions; that its space of expressions is open-ended allowing for a wide variety of possibilities to be developed; that it has no chance of finding the optimal expressions; and that it is as free from 'a priori' design restrictions as is practical and compatible with it having a bounded rationality. This agent architecture and the rationale for its structure is described in more detail in [16, 15].

4.1.3 Communication

Each agent can communicate with any of the others once a week, immediately before they all decide whether to go to the bar or not. The communication is determined by the evaluation of the talk expression and is usually either 'true' or 'false'. The presence of a quoting operator (quote) in the formal language of the talk expression allows subtrees of the talk expression to be the content of the message. If a quote node is reached in the evaluation of the talk expression then the contents of the subtree are passed down verbatim rather than evaluated. If a quoted tree is returned as the result of an evaluation of the talk expression then this is the message that is communicated.

The content of the messages can be used by agents by way of the saidBy and saidByLast nodes in the action and talk expressions. If ‘listening’ is enabled then other agents can use the message in its evaluation of its expressions – if the message is just composed of a boolean value then the saidBy node is just evaluated as this value, but if it is a more complex expression (as a result of a quote node in the sending agents talk expression) then the whole expression will be substituted instead of the saidBy (or saidByLast) node and evaluated as such. The agent can use the output of its own messages by use of other nodes (IPredictedLastWeek and ISaidYesterDay).

If ‘imitation’ is enabled then other agents can introduce any message (which is not a mere boolean value) into their own (action) gene pool, this would correspond to agents taking the message as a suggestion for an expression to determine their own action. In subsequent generation this expression can be crossed with other expressions in its population of constructs.

4.1.4 Runs of the model

Eight runs of the model were made with 10 agents in each run, each over 100 iterations. Each agent had a initial population of 40 pairs of expressions generated at random with a depth of 5.

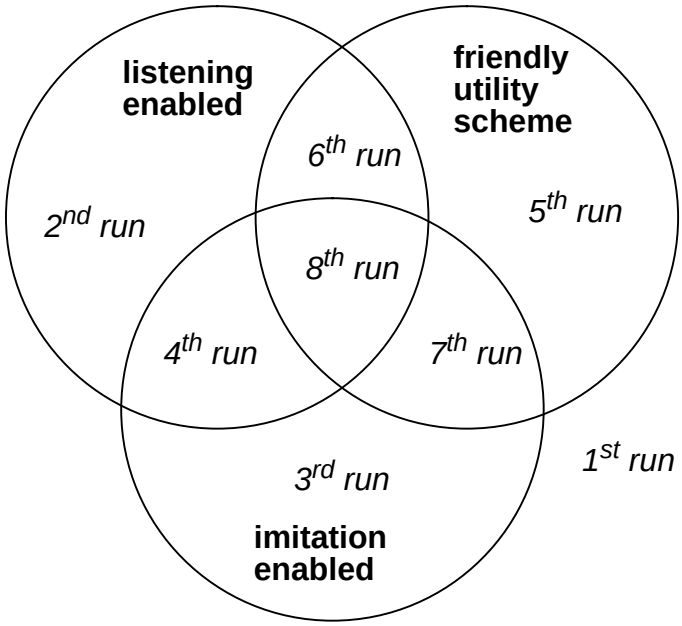

Figure 5. Variations in the 8 runs of the model

Four of the runs were done with the *friendly* scheme of expression evaluation and four with the *crowd-avoiding* scheme. In each of these clusters of four runs, in two of the runs the evaluation of saidBy and saidByLast nodes was made the same as an evaluation of a randomDecision terminal, regardless of what was actually said by the relevant agent. This had the effect of stopping agents from 'listening' to what each other said. In each pair of runs one run was with the imitation mechanism on and one was with this mechanism set as off.

In this way the eight runs cover all the combinations of: friendly/crowd-avoiding utility schemes; imitation/no imitation; listening and not listening, these possibilities are illustrated in figure 5. In this way some of the effects of these factors can be compared.

4.1.5 Implementation

The model was implemented in a language called SDML (strictly declarative modelling language), which has been developed at the Centre for Policy Modelling specifically for social modelling [26].

4.2 The Results

In figure 6 and figure 7 the attendance patterns of the agents during the eight runs are displayed. The most obvious feature is the difference between the patterns under the crowd-avoiding and friendly runs; under the crowd-avoiding scheme attendance appears far more stochastic compared to those under the friendly scheme where there is obvious coordination. This is unsurprising given that the crowd-avoiding utility scheme encourages the competitive discoordination of behaviour whilst there is a considerable advantage to (at least somewhat) coordinating action with ones 'friends' under the friendly scheme.

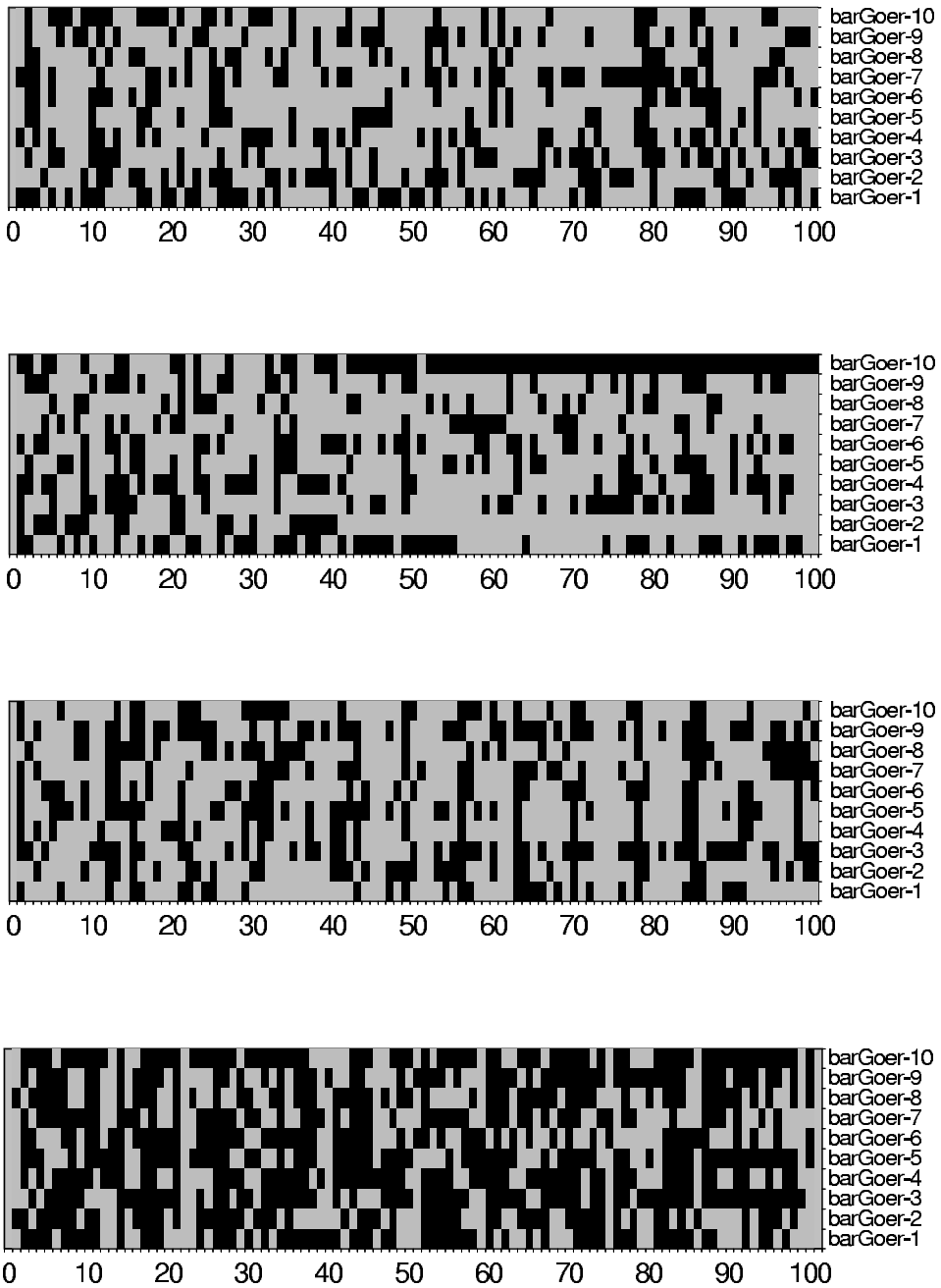

Figure 6. Attendances for the four runs under the *crowd-avoiding* scheme (grey=went, black=stayed at home)

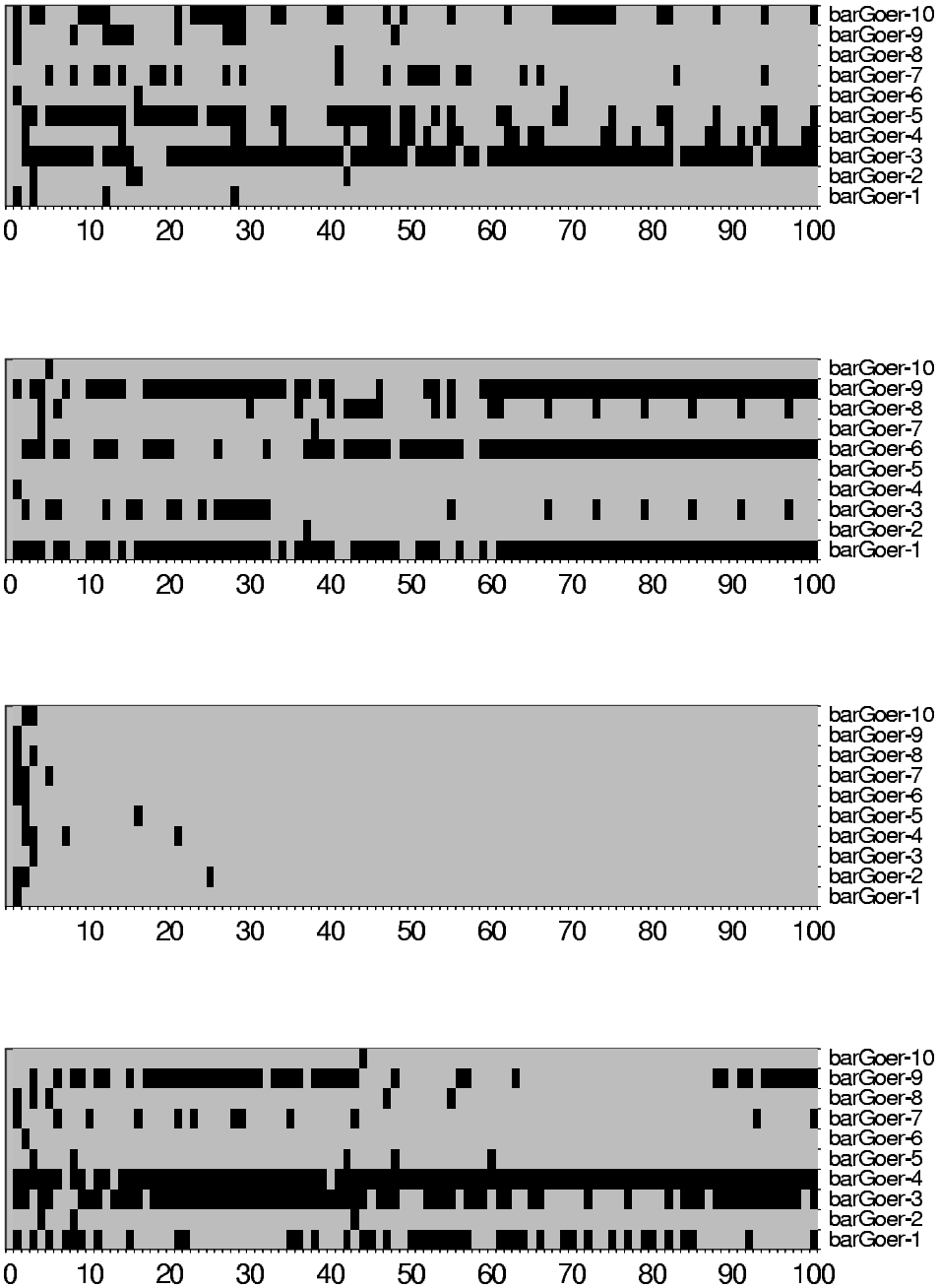

Figure 7. Attendances for the four runs under the *friendly* scheme (grey=went, black=stayed at home)

The first run exhibits the least regularity – it seems to be stochastic*. It appears that while listening and the friendly utility scheme encourage the emergence of heterogeneity among agents (i.e. there is a differentiation of strategies), imitation encourages a similarity of behaviour between agents (apparent in the vertical stripes in run 3 and the uniformity of run 7).

In table 1 and table 2, the average utility gained over the last 30 weeks and over all agents is shown for each run of the simulation. The utility gained under the *crowd-avoiding* and

friendly can not be directly compared. Under the *crowd-avoiding* scheme (table 1) it appears that both listening and imitation decrease the average utility gained, while in the runs using the *friendly* scheme (table 2) listening only had a differential impact when imitation was enabled.

	no imitation	imitation
listening	0.503	0.494
no listening	0.533	0.512

Table 1: Average utility (last 30 weeks) gained for runs under the *crowd-avoiding* scheme

	no imitation	imitation
listening	0.828	0.806
no listening	0.827	0.96

Table 2: Average utility (last 30 weeks) gained for runs under the *friendly* scheme

The next figures (figures: 8, 9, 10, 11, 12, 13, 14, and 15), show some of the specific causation between the talk and action expressions of the ten agents during the last three weeks of each run of the simulation. These figures only show the causation due the saidBy, saidByLast and wentLastWeek primitives that are active (i.e. not a saidBy or saidByLast primitive in a simulation where listening is disable and that is not logically redundant). So they do not show any causation via attendance statistics, or the self-referential primitives (e.g. ISaidYesterday, IPredictiedLastWeek and IWentLastWeek). In these figures there is a small box for the talk and action expression of each agent (numbered upwards from 1 to 10). The numbers in the boxes are a the total number of backward causal lines connected to that box if one followed the causation backward (restricted to the last three weeks only). This number is thus a indication of how socially embedded the agent is at any point in time – a larger number indicates that there is quite a complex causal chain determining the action (or

communication) of that agent, passing through many other agents. A detailed example of this (barGoer-6 in the second run) is analysed below.

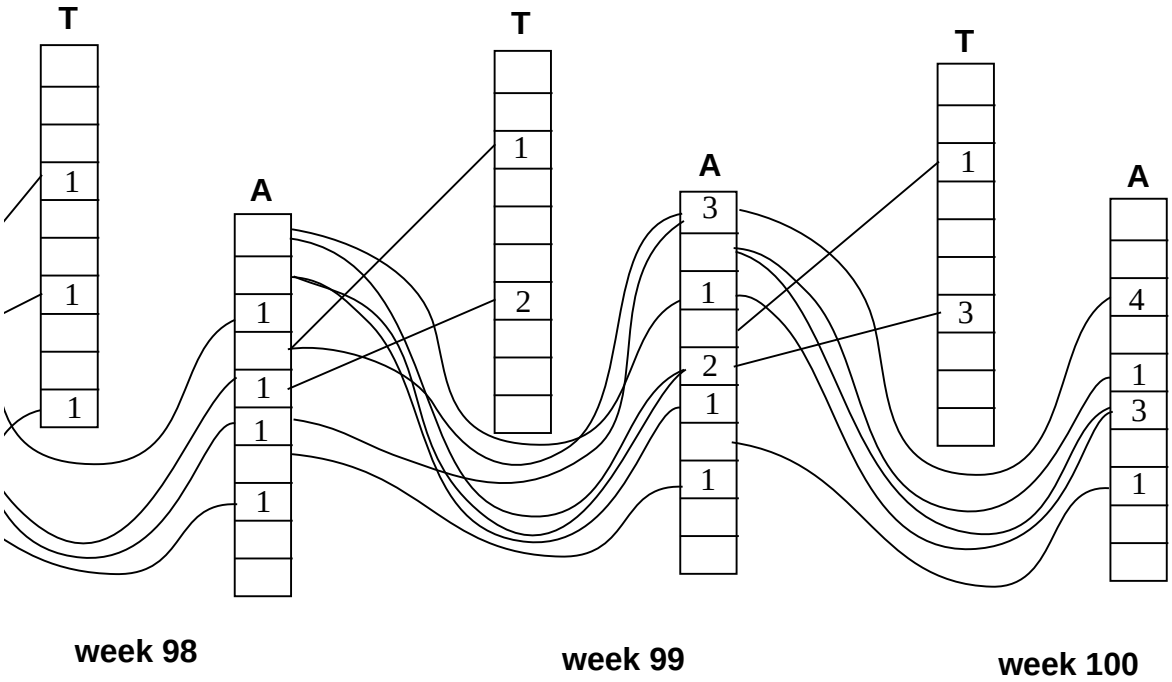

Figure 8. Causation net for run under *crowd-avoiding* scheme with neither listening nor imitation enabled

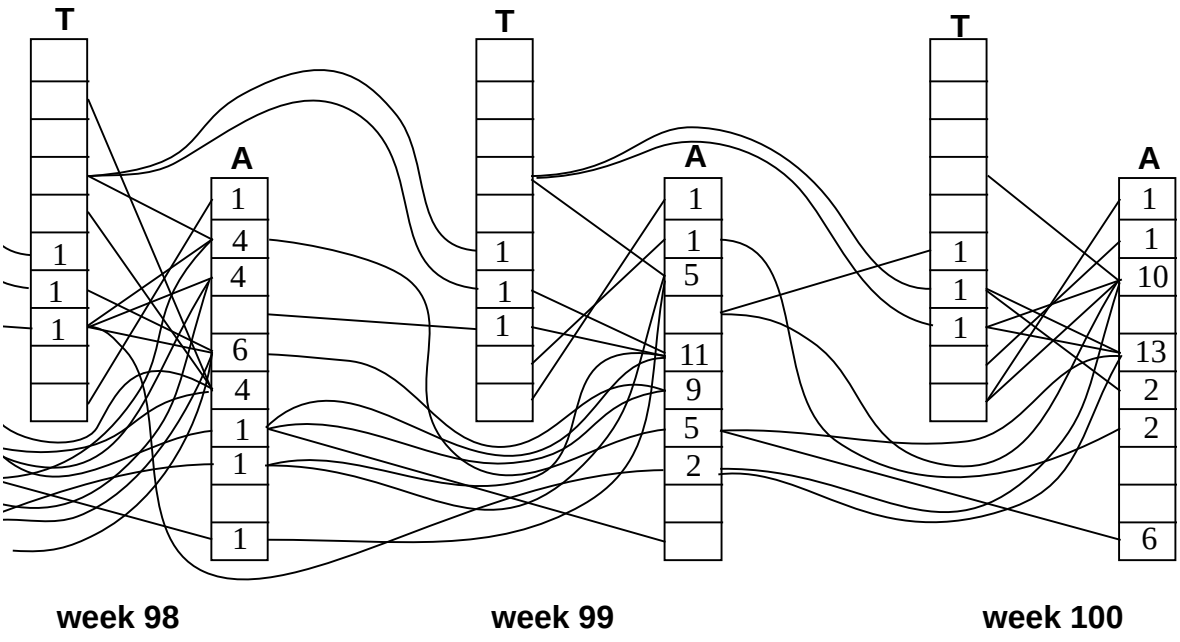

Figure 9. Causation net for run under *crowd-avoiding* scheme with only listening enabled

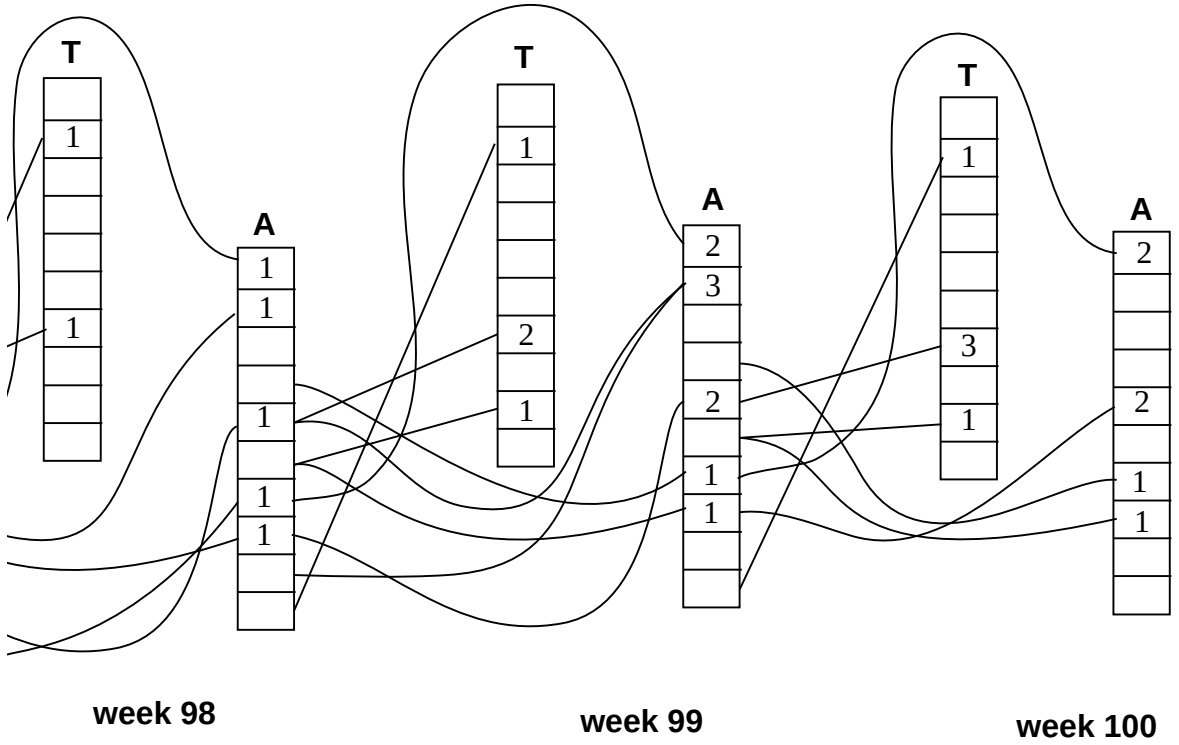

Figure 10. Causation net for run under *crowd-avoiding* scheme with only imitation enabled

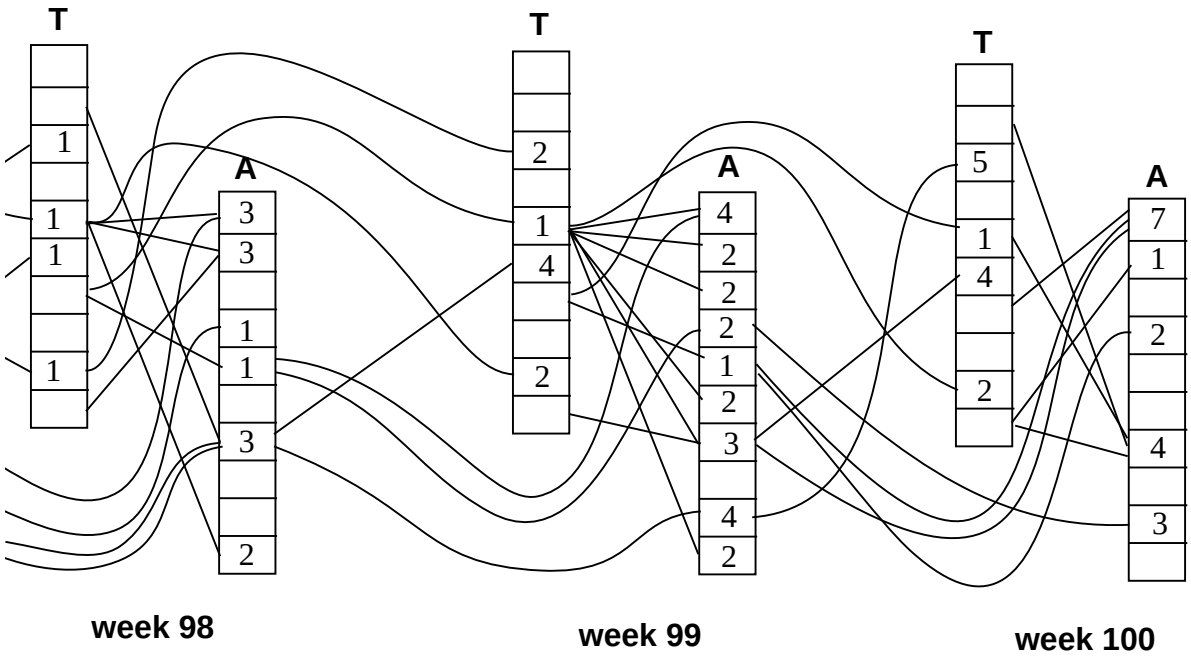

Figure 11. Causation net for run under *crowd-avoiding* scheme with both listening and imitation enabled

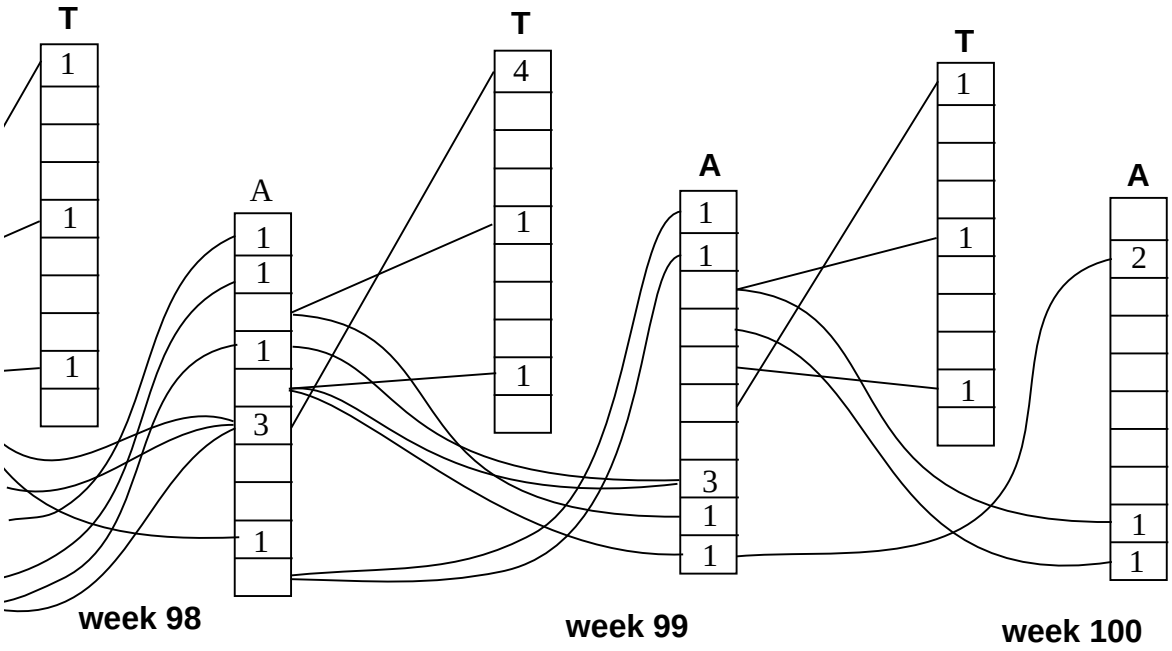

Figure 12. Causation net for run under *friendly* scheme with neither listening nor imitation enabled

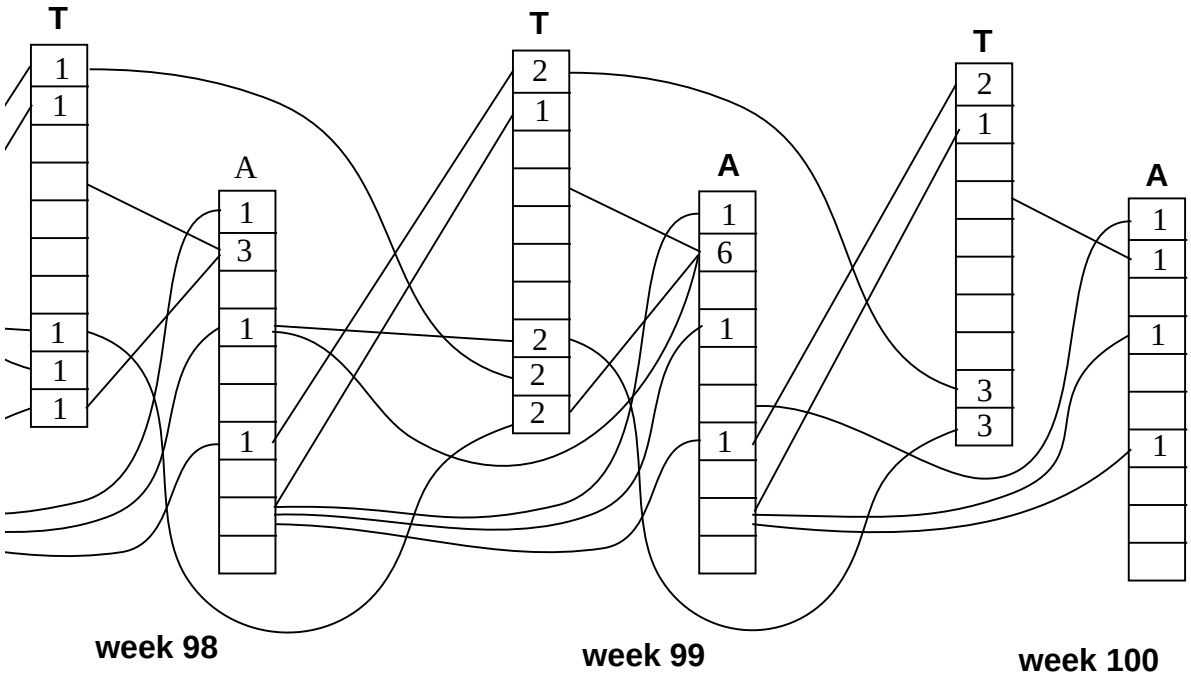

Figure 13. Causation net for run under *friendly* scheme with only listening enabled

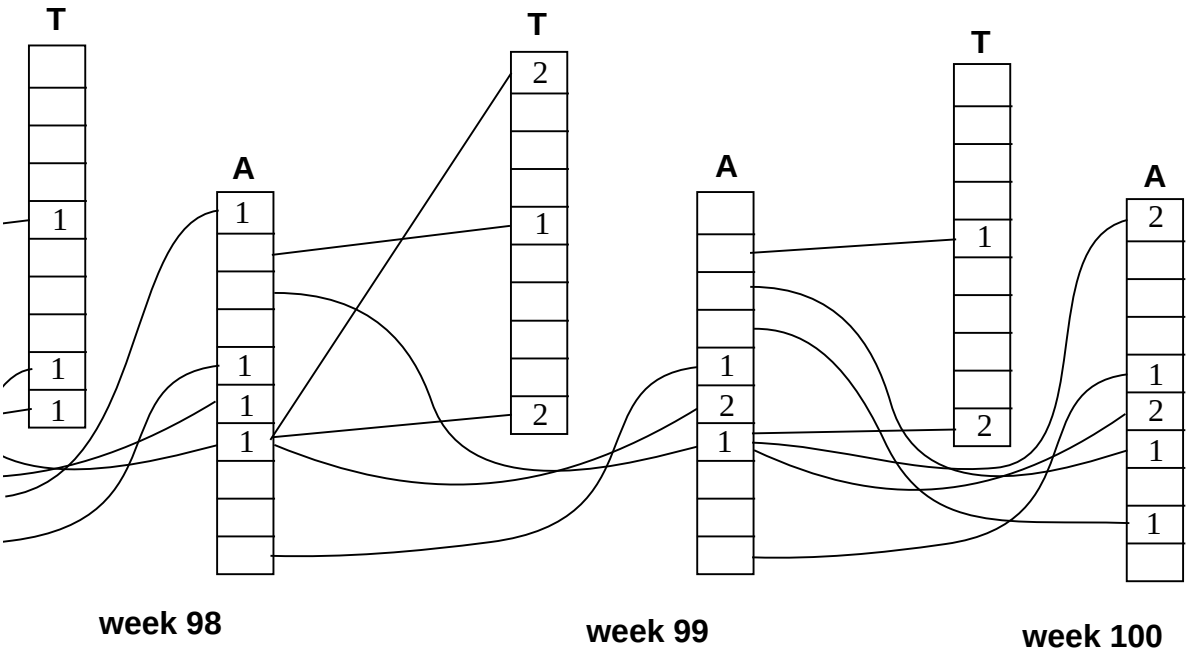

Figure 14. Causation net for run under *friendly* scheme with only imitation enabled

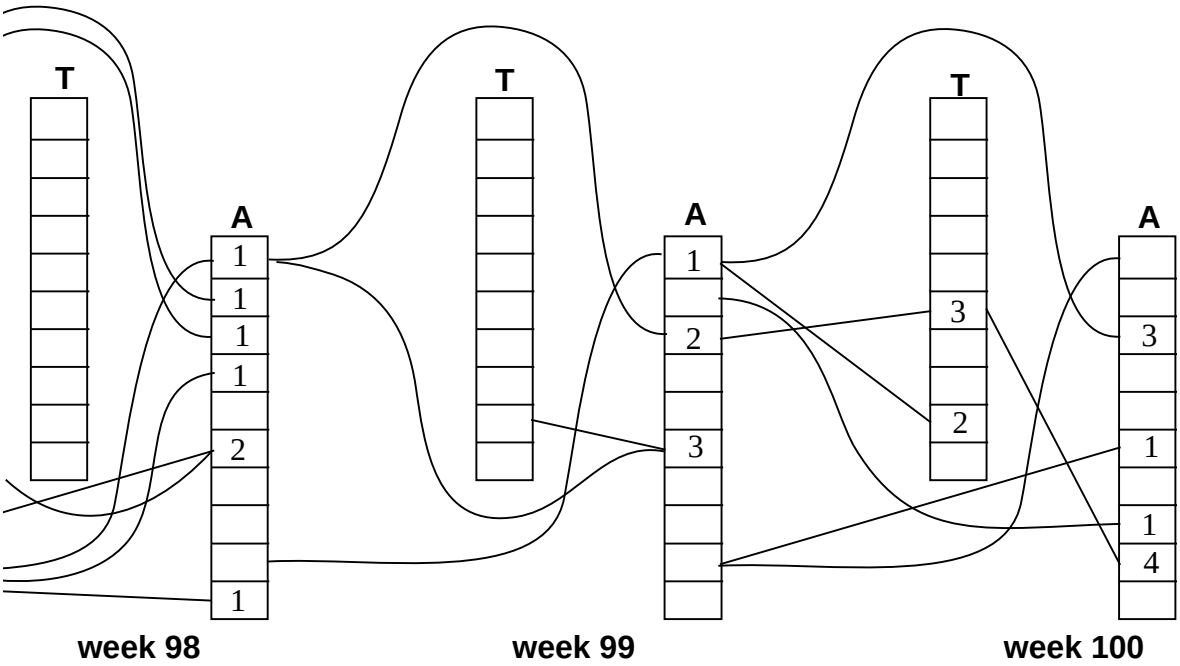

Figure 15. Causation net for run under *friendly* scheme with both listening and imitation enabled

To enable a comparison of the general levels of embedding I have tabulated the average of the last two weeks of the total of these indicators over all the agents. These numbers are shown in table 3, and table 4. These indicate that the crowd-avoiding runs of the simulation with listening enabled are more embedded that any of the other runs, with the crowd-avoiding run with listening only enabled, the most.

	no imitation	imitation
crowd-avoiding	38.5	30
friendly	15.5	10

Table 3: Embedding index for agents with listening enabled, at end of run

	no imitation	imitation
crowd-avoiding	12	12.5
friendly	9.5	9.5

Table 4: Embedding index for agents with listening disabled, at end of run

4.3 More Detailed Case Studies

In order to illustrate social embedding (or the lack of it) in these simulations, I analyse a couple of case studies of agent's behaviour and the cause one can attribute to it.

4.3.1 BarGoer-6 in the run with the *crowd-avoiding* scheme and listening only

To give a flavour of how complex a detailed explanation of behaviour can get I will follow back the chain of causation for the action of barGoer-6 at week 100.

At week 100, barGoer-6's action expression was:

```
[OR [AND [OR [AND [AND [saidBy ['barGoer-4']] [OR [AND [NOT [wentLastWeek
['barGoer-3']]] [saidBy ['barGoer-3']]] [saidBy ['barGoer-4']]]] [NOT [wentLastWeek
['barGoer-3']]]] [saidBy ['barGoer-3']]] [NOT [wentLastWeek ['barGoer-3']]]]
[wentLastWeek ['barGoer-4']]]
```

which simplifies to:

```
[OR
        [AND
                [OR
                        [saidBy ['barGoer-4']]
                        [saidBy ['barGoer-3']]]]
                [NOT [wentLastWeek ['barGoer-3']]]]
        [wentLastWeek ['barGoer-4']]]
```

substituting the talk expressions from bar goers 3 and 4 in week 100 gives:

```
[OR
        [AND
                [OR
                        [saidByLast ['barGoer-7']]
                        [wentLastWeek ['barGoer-7']]]]
                [NOT [wentLastWeek ['barGoer-3']]]]
        [wentLastWeek ['barGoer-4']]]
```

substituting the action expressions from bar goers 3, 4 and 7 in week 99 gives:

```
[OR
        [AND
                [OR
            [saidByLast ['barGoer-7']]
                [previous [OR [OR [T] [saidBy ['barGoer-2']]] [T]]]
            [NOT [previous [ISaidYesterday]]]]
        [previous [wentLastWeek ['barGoer-9']]]]
```

which simplifies to:

```
[OR
        [NOT [previous [saidBy ['barGoer-3']]]]
        [previous [wentLastWeek ['barGoer-9']]]]
```

substituting the talk expressions from barGoer-3 in week 99 gives:

```
[OR
        [NOT [previous [[wentLastWeek ['barGoer-7']]]]]
        [previous [wentLastWeek ['barGoer-9']]]]
```

substituting the action expressions from barGoers 7 an 9 in week 98 gives:

```
[OR [NOT [previous [previous [OR [OR [saidBy ['barGoer-10']] [OR [T] [OR
[randomDecision] [saidBy ['barGoer-2']]]]] [F]]]]] [previous [previous [NOT [AND
[saidBy ['barGoer-2']] [AND [AND [saidBy ['barGoer-2']] [NOT [AND [saidBy
['barGoer-6']] [wentLastWeek ['barGoer-6']]]]] [OR [AND [AND [AND [saidBy
['barGoer-2']] [OR [AND [saidBy ['barGoer-2']] [NOT [AND [saidBy ['barGoer-6']]
[wentLastWeek ['barGoer-6']]]]] [saidBy ['barGoer-2']]]] [AND [saidBy ['barGoer-2']]
[NOT [AND [AND [saidBy ['barGoer-2']] [AND [saidBy ['barGoer-2']] [saidBy
['barGoer-2']]]] [NOT [NOT [saidBy ['barGoer-2']]]]]]] [AND [randomDecision] [NOT
[saidBy ['barGoer-2']]]]]]]]
```

which simplifies to:

```
[previous [previous [NOT
[AND
        [saidBy ['barGoer-2']]
        [NOT [AND [saidBy ['barGoer-6']] [wentLastWeek ['barGoer-6']]]]]
```

substituting the talk expressions from barGoers 2 an 6 in week 98 gives:

```
[previous [previous [NOT
        [AND
                [greaterThan [1] [1]]
                [NOT [AND [[greaterThan [maxPopulation] [maxPopulation]]]
[wentLastWeek ['barGoer-6']]]]]
```

which simplifies to:

```
True
```

The above trace ignores the several important causal factors: it does not show the evolutionary processes that produce the action and talk genes for each agent at each week; it does not show the interplay of the agent's actions and communications upon events and hence the evaluation of expressions (and hence which is chosen next by all agents); and in simplifying the expressions at each stage I have tacitly ignored the potential effects of the parts of the expressions that are logically redundant under this particular train of events. Even given these caveats the action of barGoer-6 at week 100 was determined by a total of 11 expressions: its choice of the action expression shown; the talk expressions from bar goers 3 and 4 in week 100; the action expressions from bar goers 3, 4 and 7 in week 99; the talk expressions from barGoer-3 in week 99; the action expressions from barGoers 7 an 9 in week 98; and the talk expressions from barGoers 2 an 6 in week 98!

On the other hand it is difficult to find models of the behaviour of barGoer-6 which does not involve the complex web of causation that occurs between the agents. It is not simplistically dependent on other particular agents (with or without a time lag) but on the other hand is not merely random. This agent epitomises, in a reasonably demonstrable way, social embeddedness.

4.3.2 BarGoer-9 in the run with the *friendly* scheme and listening only

At week 100 the selected talk and action expressions for the 10 agents were as below.

barGoer-3's (talk) [wentLastWeek ['barGoer-7']]
barGoer-3's (action) [OR [OR [OR [friendOfMine ['barGoer-2']] [friendOfMine ['barGoer-2']]] [friendOfMine ['barGoer-5']]] [friendOfMine ['barGoer-1']]]
barGoer-6's (talk) [lessThan [numWentLastTime] [numWentLastTime]]
barGoer-6's (action) [friendOfMine ['barGoer-2']]
barGoer-7's (talk) [greaterThan [10] [10]]
barGoer-7's (action) [wentLastWeek ['barGoer-2']]
barGoer-4's (talk) [greaterThan [3] [3]]
barGoer-4's (action) [wentLastWeek ['barGoer-2']]
barGoer-1's (talk) [saidByLast ['barGoer-3']]
barGoer-1's (action) [AND [AND [saidBy ['barGoer-8']] [AND [wentLastWeek ['barGoer-2']] [wentLastWeek ['barGoer-8']]]] [AND [AND [saidBy ['barGoer-8']] [AND [NOT [wentLastWeek ['barGoer-8']]] [AND [wentLastWeek ['barGoer-8']] [AND [saidBy ['barGoer-8']] [AND [wentLastWeek ['barGoer-2']] [AND [T] [AND [wentLastWeek ['barGoer-8']] [AND [saidBy ['barGoer-4']] [AND [saidBy ['barGoer-8']] [AND [wentLastWeek ['barGoer-6']] [wentLastWeek ['barGoer-8']]]]]]]]]]] [AND [wentLastWeek ['barGoer-8']] [AND [wentLastWeek ['barGoer-8']] [AND [AND [AND [saidBy ['barGoer-6']] [AND [AND [AND [saidBy ['barGoer-8']] [AND [NOT [wentLastWeek ['barGoer-8']]] [AND [wentLastWeek ['barGoer-8']] [AND [wentLastWeek ['barGoer-8']] [AND [saidBy ['barGoer-8']] [saidBy ['barGoer-8']]]]]]] [AND [wentLastWeek ['barGoer-6']] [AND [NOT [wentLastWeek ['barGoer-6']]] [AND [wentLastWeek ['barGoer-6']] [AND [AND [wentLastWeek ['barGoer-8']] [AND [wentLastWeek ['barGoer-8']] [AND [saidBy ['barGoer-8']] [AND [T] [T]]]]] [wentLastWeek ['barGoer-8']]]]]] [wentLastWeek ['barGoer-8']]]] [wentLastWeek ['barGoer-8']]] [AND [saidBy ['barGoer-8']] [wentLastWeek ['barGoer-8']]]]]]]
barGoer-8's (talk) [friendOfMine ['barGoer-4']]
barGoer-8's (action) [OR [NOT [NOT [NOT [NOT [friendOfMine ['barGoer-9']]]]]] [NOT [friendOfMine ['barGoer-9']]]]
barGoer-9's (talk) [wentLastWeek ['barGoer-2']]
barGoer-9's (action) [AND [saidBy ['barGoer-7']] [wentLastWeek ['barGoer-1']]]
barGoer-10's (talk) [wentLastWeek ['barGoer-4']]
barGoer-10's (action) [wentLastWeek ['barGoer-5']]
barGoer-5's (talk) [lessThan [7] [7]]
barGoer-5's (action) [friendOfMine ['barGoer-2']]
barGoer-2's (talk) [saidByLast ['barGoer-10']]
barGoer-2's (action) [friendOfMine ['barGoer-5']]

Although many of these are simply reducible to True or False, others are not. Further more, although many of these expressions remained pretty much constant over the last 10

weeks of the simulation some did not. For example the action expressions of barGoer-9 during the last 10 weeks were:

91: [AND [AND [wentLastWeek ['barGoer-1']] [AND [AND [AND [AND [saidBy ['barGoer-1']] [wentLastWeek ['barGoer-7']]] [wentLastWeek ['barGoer-7']]] [saidBy ['barGoer-7']]] [wentLastWeek ['barGoer-7']]]] [saidBy ['barGoer-1']]]
92: [AND [wentLastWeek ['barGoer-7']] [saidBy ['barGoer-7']]]
93: [AND [AND [wentLastWeek ['barGoer-7']] [AND [wentLastWeek ['barGoer-7']] [AND [saidBy ['barGoer-1']] [saidBy ['barGoer-7']]]]] [saidBy ['barGoer-1']]]
94: [AND [AND [wentLastWeek ['barGoer-7']] [AND [wentLastWeek ['barGoer-7']] [AND [saidBy ['barGoer-1']] [saidBy ['barGoer-7']]]]] [saidBy ['barGoer-1']]]
95: [AND [AND [saidBy ['barGoer-1']] [wentLastWeek ['barGoer-1']]] [AND [wentLastWeek ['barGoer-7']] [saidBy ['barGoer-1']]]]
barGoer-10's uses talk gene [wentLastWeek ['barGoer-4']]
96: [saidBy ['barGoer-7']]
97: [AND [wentLastWeek ['barGoer-7']] [AND [wentLastWeek ['barGoer-7']] [AND [saidBy ['barGoer-1']] [saidBy ['barGoer-7']]]]]
98: [AND [saidBy ['barGoer-7']] [saidBy ['barGoer-1']]]
99: [AND [wentLastWeek ['barGoer-7']] [AND [wentLastWeek ['barGoer-7']] [AND [saidBy ['barGoer-1']] [saidBy ['barGoer-7']]]]]
100: [AND [saidBy ['barGoer-7']] [wentLastWeek ['barGoer-1']]]

Each time barGoer-9's action expression is a conjunction of saidBy or wentLastWeek referring to agents barGoer-1 and barGoer-7. Each time [wentLastWeek ['barGoer-1']] and [saidBy ['barGoer-7']] would evaluate to False and [wentLastWeek ['barGoer-7']] and [saidBy ['barGoer-1']] to True, so its continued non-attendance would depend upon the presence of either of a [wentLastWeek ['barGoer-7']] or [saidBy ['barGoer-1']] in the chosen conjunction.

But in this run of the simulation there is a far simpler explanation for bar-Goer-9's behaviour: that is because it has only two 'friends' (barGoer-10 and barGoer-3) it is not worth its while to attend. In fact this is true for each agent – its attendance pattern can be explained almost entirely on the number of friends it has (figure 2 shows the imposed friendship structure for this run). This is shown in table 5. Only bar goers 3, 8 and 7 need further explanation. BarGoer-7 has three friends but none of these are 'loners' like barGoer-9 (i.e. only having 2 friends), so there is a good chance that three of its friends will go while barGoer-3 and 8 both have a friend who is a loner. The behaviour with period 6 arises because agents evaluate their expressions only up to five time periods ago, and so every sixth week barGoer-3 and 8 have 'forgotten' their previous (unsuccessful) attendance and go again*.

Thus in this case we have a simple explanation of barGoer-9's continued absence from the bar in terms of its own likely utility due to the limited number of friends it has*. Agents barGoer-3 and 8 are slightly more embedded that the others at the end of this run as the explanation of their behaviour has to include each other and the fact that they have friends who only have two friends.

4.4 Comments

The simulation exhibits most of the effects listed above (in the section previous to the description of the model set-up). This is, of course, unsurprising since I have been using the model to hone my intuitions on the topic; the ideas about social embeddedness and the model have themselves co-developed. In particular:

- the expressions that the agents develop resemble constructs rather than models, in that they are opportunistic, they do not reflect their social reality but rather constitute it;
- the constructs can appear highly arbitrary – it can take a great deal of work to unravel them if one attempts to explicitly trace the complex networks of causation (see the examples in the case studies above);
- the agents do frequently use information about the communication and actions of others in stead of attempting to explicitly predict their environment – this is partly confirmed by a general analysis of the general distribution of primitive types in the expressions chosen and developed by agents in figure 16 (the categories the primitives are collected into are fairly self explanatory);
- the agents do specialise as they co-develop their strategies – this is not so apparent from the above but is examined in greater depth elsewhere [14];

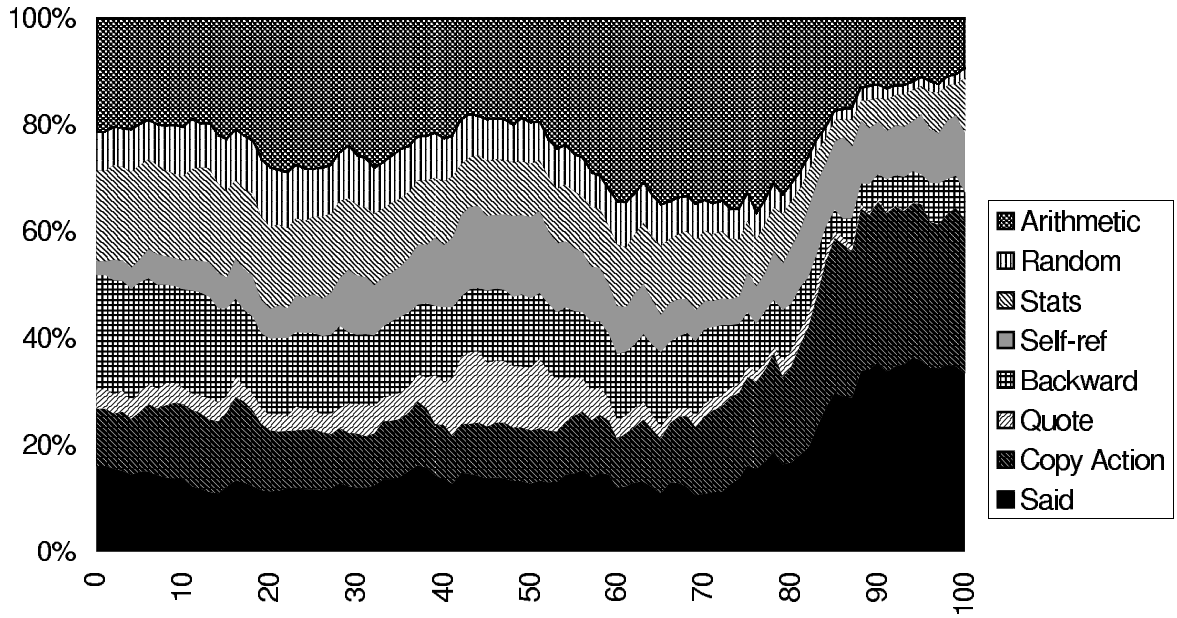

Figure 16. Distribution of the relative proportions of some primitive types in the run using the *crowd-avoiding* scheme with only listening enabled

It is not obvious in the runs but does *seem* to be the case that a Wittgensteinian analysis of the language use by the agents might be appropriate. It did seem to be in a simulation similar to the crowd-avoiding run with only listening enabled described in [15]. It was also unclear whether there was anything that might correspond to the emergence of social structures, but I would expect that such would only result from longer and more sophisticated simulations than the above.

5 Conditions for the Occurrence of Social Embedding

What might enable the emergence of social embeddedness? At this point one can only speculate, but some factors are suggested by the above model. They might be:

- the ability of agents significantly to ef fect their environment – so that they are not limited to an essentially passive predictive role;
- the co-development of the agents – for example, if agents had co-evolved during a substantial part of the development of their genes then it is likely that this evolution would have taken advantage of the behaviour of the other agents; this would be analogous to the way dif ferent mechanisms in one or ganism develop so that they have multiple and overlapping functions that defy their strict separation [30];
- the existence of exploitable computational resources in the environment (in particular , the society) – so that it would be in the interest of agents to use these resources as opposed to performing the inferences and modelling themselves;
- the possibility of open-ended development by the agents – if the space of possible constructs was essentially small (so that an approximation to a global search could be

preformed), then the optimal model of the society that the agent inhabited would be feasible for it;

- mechanisms for social distinction, hence implicit (or explicit) modelling, relationships
- the ability to develop the *selection* of information sources – which depends on there being a real variety of distinguishable sources to select from;
- the ability to frequently sample and probe social information (i.e. gossip), thus our intelligence might both have enabled the development of social embedding as well as being selected for it (as in the 'social intelligence hypothesis' of [22]).

What is very unclear from the above model and analysis is the role that imitation plays in the development (or surpression) of social embeddedness, particularly where both imitation and conversational communication are present. In [8], Kerstin suggests that imitation may have a role in the effectiveness of an agent to cope with a complex social situation (or rather not cope as a result of autism). The above model suggests imitation may have a role in simplifying social situations so that such embedding is unnecessary.

6 Conclusion

Despite the fact that I have characterised social embedding in a constructivist way, its presence can have real consequences for any meaningful models of social agents that we create. It is not simplistically linked to coordination, communication or motivation but may interact with these.

Probably its application will have the most immediate impact upon our modelling methodologies. For example, it may help to distinguish which of several modelling methodologies are most useful for specified goals. It might be applied to the engineering of agent communities so as to help reduce unforeseen outcomes by *surpressing* social embedding. Hopefully social embeddedness can be identified and analysed in a greater variety of contexts, so as to present a clearer picture of its place in the modelling of social agents.

Acknowledgements

Thanks to Scott Moss, Helen Gaylard for many discussions, to Steve Wallis for writing SDML and to Kerstin Dautenhahn for organising the Socially Intelligent Agents workshop in Boston in November 1997, which stimulated the production of this paper.

SDML has been developed in VisualWorks 2.5.1, the Smalltalk-80 environment produced by ObjectShare (formerly ParcPlace-Digitalk). 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] Akiyama, E. and K. Kaneko, (1996). Evolution of Cooperation, Differentiation, Complexity, and Diversity in an Iterated Three-person Game, *Artificial Life*, 2:293-304.

[2] Arthur, B. (1994). Inductive Reasoning and Bounded Rationality. *American Economic Association Papers*, 84: 406-411.

[3] Bednarz, J. 1984. Complexity and Intersubjectivity. *Human Studies*, 7:55-70.

[4] Beer, R. D. (1990). *Intelligence as Adaptive Behaviour.* Academic Press.

[5] Brooks, R. (1991). Intelligence without Representation. *Artificial Intelligence*, 47:139-159.

[6] Carley, K. and Newell, A. (1994). The Nature of the Social Agent. *Journal of Mathematical Sociology*, 19:221-262.

[7] Carneiro, R. L. (1987). The Evolution of Complexity in Human Societies and its Mathematical Expression. *International Journal of Comparative Sociology*, 28:111-128.

[8] Dautenhahn, K. (1997). I Could Be You: The phenomenological dimension of social understanding. *Cybernetics and Systems*, 28:417-453.

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

[10] Durkheim, E. (1895). *The rules of sociological method*. Readings from Durkheim. chichester: Ellis Horwood.

[11] Edmonds, B. (1996). Pragmatic Holism. CPM Report 96-08, MMU, 1996.

[12] Edmonds, B. (1997). Complexity and Scientific Modelling. 20th International Wittgenstein Symposium, Kirchberg am Wechsel, Austria, August 1997. Also available as a CPM Report 97-23, MMU, Manchester, UK.

[13] Edmonds, B. What is Complexity?: the philosophy of Complexity per se with application to some examples in evolution. In F. Heylighen & D. Aerts (eds.): *The Evolution of Complexity*, Dordrecht: Kluwer.

[14] Edmonds, B. (1998). Modelling Bounded Rationality In Agent-Based Simulations using the Evolution of Mental Models.Workshop on Agent-based and Population-based Modelling of Learning in Economics, Max-Planck Institute for Research into Economic Systems, Jena, Germany, March 1998.

[15] Edmonds, B. (forthcoming). Modelling Socially Intelligent Agents. *Applied Artificial Intelligence*.

[16] Edmonds, B. and S. J. Moss, (1997). Modelling Bounded Rationality using Evolutionary Techniques. *Lecture Notes in Computer Science*, 1305:31-42.

[17] Foerster, E. von (1973). On Constructing a Reality. In Preiser (ed.), Environmental Research Design, Vol 2. Stroudsburg: Dowden, Hutchinson and Ross, 35-46.

[18] Franklin, S. (1996). Coordination without Communication. University of Memphis, 1996.

[19] Glasersfeld, E. von (1995). Radical Constructivism: A Way of Knowing and Learning. London: the Falmer Press.

[20] Grassé P. P. (1959). La reconstruction du nid et les coordinations inter-individuelles chez Bellicositermes natalensis et Cubitermes sp. La theorie de la stigmergie: Essai d'interpretation des termites constructeurs. *Insect Societies*, 6:41-83.

[21] Koza, J.R. (1992). *Genetic Programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.

[22] Kummer, H., Daston, L., Gigerenzer, G. and Silk, J. (1997). The social intelligence hypothesis. In Weingart et. al (eds.), *Human by Nature: between biology and the social sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, 157-179.

[23] Montana, D. J. (1995). Strongly Typed Genetic Programming, *Evolutionary Computation*, 3:199-230.

[24] Moss, S.J. and Edmonds, B. (1998). Modelling Economic Learning as Modelling. *Systems and Cybernetics*, 29:5-37.

[25] Moss, S. J., Edmonds, B. and Wallis, S. (1997). Validation and Verification of Computational Models with Multiple Cognitive Agents. CPM Report 97-25, MMU, 1997.

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

[27] Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Ballentine.

[28] Riegler, A. (1992). Constructivist Artificial Life and Beyond. Workshop on Autopoiesis and Perception, Dublin City University, Aug. 1992.

[29] Vaario, J. (1994). Artificial Life as Constructivist AI. *Journal of SICE*.

[30] Wimsatt, W. (1972). Complexity and Organisation, in Scavenger and Cohen (eds.), Studies in the Philosophy of Sciences. Dordrecht: Riddle, 67-86.

[31] Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.