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. Social embeddedness is intended as a strong type of social situatedness. It is defined as the extent to which modelling the behaviour of an agent requires the inclusion of other agents as individuals rather than as an undifferentiated whole. Possible consequences of the presence of social embedding and ways to check for it are discussed. A model of co-developing agents is exhibited which demonstrates the possibility of social embedding. This is an extension of Brian Arthur's 'El Farol Bar' model, with added learning and communication. Some indicators of social embedding are analysed and some possible causes of social embedding are discussed. It is suggested that social embeddedness may be an explanation of the causal link between the social situatedness of the agent and it employing a constructivist strategy in its modelling.

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 one way in which the *social* situation can effect an agent. It aims to identify phenomena that may be usefully taken to indicate the extent to which agents are inextricably embedded in a society of other such agents. In particular, it aims to show this for a particular artificial simulation involving co-evolving agents. In order to do this a modelling approach is adopted which takes ideas from 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 computational 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 has to be 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 [28] to denote the process whereby an individual constructs its view of the world. Extrapolating from this is Ernst von Glasersfeld's 'radical constructivism' [18] 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).

There is not enough room to survey this rich philosophical position. So for the purposes of this paper I will list some the aspects of constructivism that are relevant for my purposes here*:

- models† that an agent builds do not necessarily *reflect* the structure of agent's environment (as viewed by an external observer) – rather the models are merely compatible with the environment and the agent's interactions with that environment;
- models are developed with respect to the needs and goals of the agent, particularly with respect to its attempts to control its own actions and that of its environment;
- models are built up as a result of active interaction with its environment rather than as a result of passive observation and reasoning – in fact, the models may well require interaction with the environment in order to function as action selection mechanisms;
- it emphasises the bottom-up approach to modelling, with a tendency away from *a priori* considerations;

Constructivism has been taken up by some researchers in artificial intelligence and artificial life (e.g. [10, 29, 31]) 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 it is the serendipidous exploitation of features of its environment and the strong practical interaction *during* development which makes it effective. 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 approach which subsumes the work of Rodney Brooks [5], but instead of the development of the organism happening through an analysis, design and test cycle done by human designers based on their knowledge, 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 models of the systems we are investigating, rather than some aspect of an external independent reality, because I claim that a *useful* characterisation of social embeddedness has to take into account the modelling framework.
- *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 the

* For those who want to know more about the wider framework that is constructivism, a good introduction from a philosophical and cybernetic perspective can be found at the Principia Cybernetica web site at URL: http://pespmc1.vub.ac.be/construc.html

† Some authors distinguish between models and constructs – the implication being that models are reflections of the external environment. No such connotation is intended here, in this paper I use 'model' in a weak sense, it could be merely be the machinery that singles out the action given the agent's state and perceptions/inputs.

internal models are grounded in their effect upon the agent in conjunction with other agent's actions. The *meaning* of the agent's communication is not fixed beforehand by the programmer, so the effect of such communication and action is grounded [20] in its use in practice and in the language-games that the agents appear to play [34].

- *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 explanation for other social situations.

3 Characterising Social Embeddedness

In attempting to elucidate the concepts of 'social situatedness' or '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, when he claims that some social phenomena should be considered *entirely separately* from individual phenomena [11]. Cognitive science has the opposite perspective – the individual's behaviour and processes are primitive and the social phenomena may emerge as an emergent *result* of such individuals interacting.

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

For this reason I will take a pragmatist approach and suggest the categorisation of social systems relative to some pertinent modelling considerations. This is based on a philosophy of *pragmatic holism* which is constructivist in style. Its essence is that regardless of whether the natural world *is* theoretically reducible we have to *act* as if there are irreducible wholes. This means that we should explicitly include aspects of the modelling process in our theories. For more on this position see [12]. Thus, 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. I want to avoid the idealisations involved in this disputed area and concentrate on what can are *useful* attributions in describing social situations and their computational analogs.

3.1 Being Situated

When Brooks [5] made his now famous critique of AI (as it was then). He was specifically addressing shortcomings with respect to the problem of getting robots to master a *physical* environment. This spawned a whole field of research based on the premiss that the physical situation was critically important in the design of agents (and in particular robots).

Since then the property of 'being situated' has been characterised in many (subtly different) ways. For example, Alonso Vera and Herbert Simon [32] argue that the characteristics of situated action is the utilisation of external rather than internal representations via the functional modelling of the affordances provided by the environment. In their account this allows the paring down of the internal representation so that its processing can occur in real-time.

More recently William Clancey, in attempting to forge some sort of consensus on the subject wrote (page 344 of [8]):

> *"In summary, the term situated emphasises that perceptual-motor feedback mechanisms causally relate animal cognition to the environment and action in a way that a mechanism based on logical (deductive inference alone does not capture."*

What these various approach agree upon is that if you are to effectively model certain domains of action over time then you need to include sufficient detail of the environment so that explanations of choice of action can be made in terms of the detailed causal chains via this environment. In other words, the actions will not be satisfactorily explained with reference to internal inference processes alone, but only by including causal feedback from the environment.

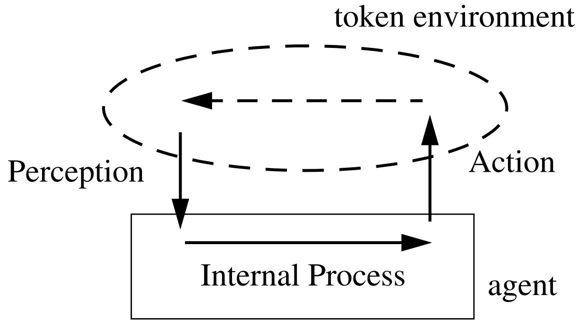

Figure 1. Where the internal inference is a sufficient as the model for action

This can be summarised (a little crudely) by saying that in a non-situated agent the internal 'inferential' processes form a sufficient model for the relationship between perception and action (figure 1), whereas when an agent is situated you need to also include the exterior causation to form a sufficient model of this relationship (figure 2). Of course, if the agent was making a one-shot decision the pictures would be equivalent in effect since the causal part of the loop would not be needed in determining the relationship between perception and action, but more usually the loop is traversed many times, with several past actions and perceptions, in order to determine the next action.

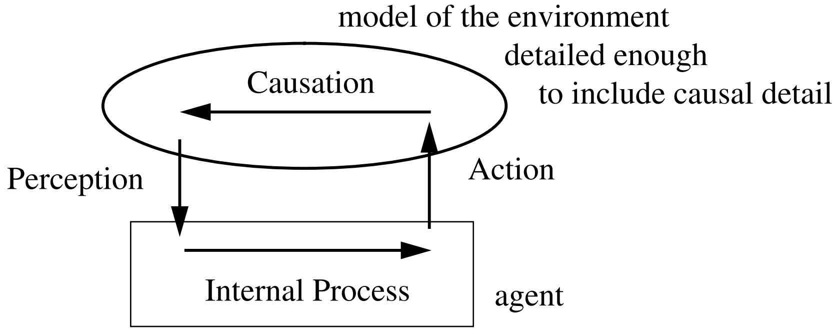

Figure 2. Where external causation is also part of the model for action

Being situated has practical considerations for what might be effective decision strategies on behalf of the agent. If internal models alone are likely to be insufficient (or just too difficult), and there are implicit computational and representational resources in the environment it make sense to make use of these by 'probing' them frequently for information as to effective action. This fits in with Lucy Suchman's characterisation of situatedness which is as follows (page 179 of [30]):

> *"... the contingence of action on a complex world ... is no longer treated as an extraneous problem with which the individual actor must contend, but rather is seen as an essential resource that makes knowledge possible and gives action its sense. ... the coherence of action is not adequately explained by either preconceived cognitive schema or institutionalised social norms. Rather the organisation of situated action is an emergent property of moment-by-moment interactions ..."*

3.2 Being Socially Situated

In a physical situation the internal models may be insufficient because of the enormous computation capacity, amount of information and speed that would be required by an agent attempting to explicitly model its environment. In a social situation, although the speed is not so critical, the complexity of that environment can be overwhelming and there is also the obvious external computational resources provided by the other agents and their interactions. This means that an agent can be said to be socially situated by analogy with being physically situated – in both cases the balance of advantage lies in using external causal processes and representations rather than internal ones. The fact that the source of this imbalance in each case is due to different causes leads to a different 'flavour' of the situatedness, but there is enough in common to justify the common use of word 'situated'. Of course, social environments vary greatly and the fact of being socially situated will thus be contingent on the particular agent and its social context.

The frequent sensing and probing of the physical environment can be translated into 'gossip', one of whose functions is the frequent sampling and testing of the social environment. The reliance of external computational resources and models is arguably even more pronounced in social situations than physical ones – social agents may accept the output of external sources (including other agents) as a direct influence on their decision making, e.g. in fashion.

3.3 Being Socially Embedded

Extending the above characterisations of situatedness, 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 saying an agent is socially embedded is stronger than saying it is merely socially situated. I have characterised social embeddedness as a *construct* which depends on one's modelling goals, since these goals will affect the criteria for the appropriateness of models. It can be read as contrasting 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 3.

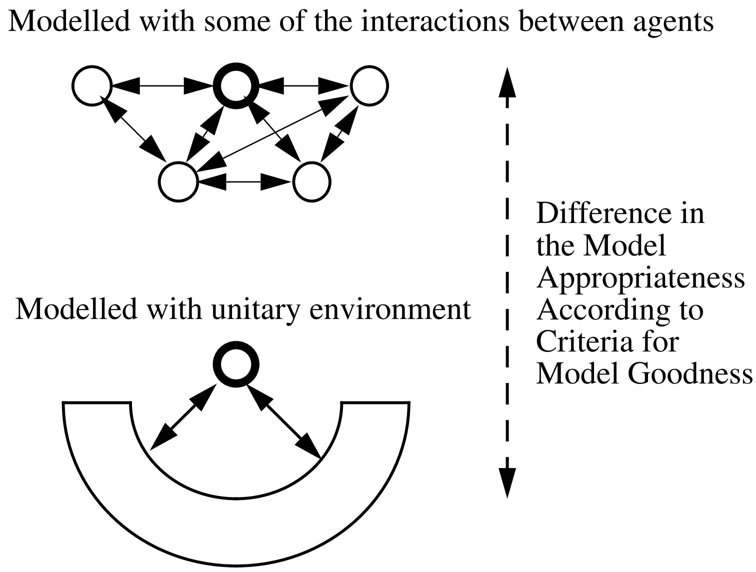

Figure 3. Social embeddedness as the appropriate level of modelling

This is not an extreme ‘relativist’ position since, if one fixes the modelling framework and criteria for model selection, the social embedding of agents within a collection of agents can sometimes be unambiguously assessed. When the modelling framework is agreed, the object of modelling (in this case ‘social systems’) will constrain the models that fit the framework. If one is extremely careful (and lucky) it might entail a unique model – in such cases we can safely project the social embeddedness upon the social system itself. Note however, that this projective attribution onto the social system is a *post-hoc* attribution that can only occur unambiguously in special circumstances. Usually there will be many arbitrary choices involved in the modelling of the social phenomena, so that the model (and hence the social embeddedness) is underdetermined by the phenomena itself. It is for this reason that it is more useful to define the social embeddedness with respect to model properties and use the association of the best model (by the chosen model selection criteria) with the phenomena itself as a means of inferring properties on the object system.

According to this account the social embedding is dependent on the modelling framework. Such a modelling framework includes the language of model representation, the model selection criteria and the goals of modelling. Frequently such a framework is implicitly agreed but not always. I have not the space here to fully specify what such a framework entails, for more details on this see [13, 25].

Notice that criteria for model acceptability can include many things other than just its predictive accuracy, for example: *complexity* [13]. 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) separate 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 consider some examples to make this a little clearer.

- Firstly a neo-classical 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 representative 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.
- Secondly where 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.
- Thirdly, the behaviour of a 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 stigmergy: [19]). Here one could say that the ants were socially *situated* but not socially *embedded*, since one can model the system with an essentially unitary model of the environment, which each of the ants separately interact with.
- Finally, in modelling the movements of people at a party, it is possible that to get any reasonably accurate results one would have to include explicit representations of each person and their relationship with each of the others present. This would represent a high level of social embeddedness.

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 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.). So it is not enough that the outcomes of the model are verified and the structure validated (as in [26]) 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 interacting 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. 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.

* A similar condition is that the agents should be essentially homogeneous.

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*.

3.4 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 is twofold: firstly, it will be to its advantage (in general) to include individual specific elements in its internal decision making processes and secondly, a complete model of its environment will be impossible. 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 models, where these are grounded in possible action*;
- as a result the models of an agent may appear somewhat arbitrary (to an external observer);
- it is worth frequently sampling and interactively testing its social environment to stand instead of complete internal models of that environment (e.g. engage in gossip);
- agents specialise to inhabit a particular social niche, where some subset 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 summarised 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).

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.

3.5 Checking for Social Embeddedness

Given that 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 described above to specify the social embeddedness, 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 [25]. 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

* This may be moderated by the riskiness of the actions involved.

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 power might be, it is easier to proceed by making judgements as to the complexity of models.

In the model below I 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 subexpressions in their learnt strategies which preform an external reference to the individual actions of other agents to those 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

The model described below is illustrative – it illustrates the *possibility* of social embedding. Despite the obvious *analogies* with human social interaction, it does not attempt to be descriptively realistic. Instead it is designed to reveal the sort of phenomena that *can* emerge in a collection of socially situated agents – ones that co-develop behavioural strategies in an open-ended way, where these strategies can include references to the actions and utterances of specific agents in their society. The choice of the learning algorithm based on genetic programming is so as to bias the agents as little as possible with *a priori* specifications of the 'desirable' strategy, but to allow the emergence of behaviours from their reaction to the environment and each other. The primitives that determine the range of strategies is designed to be as expressive as possible, thus it allows everything from purely social strategies such as following a leader, to implementations of the sort of randomised mixed strategy that might be suggested by game theory.

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. 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 numeric utility, but this only impacts on the model in the agent's evaluations of their models. 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: *crowd-avoiding* and *friendly.* The first of these encourages the straightforward discoordination of the agents actions and the second is a mixture of discoordination and cooperation. The contrast between them is designed to bring out the extent to which embedding may be effected by the motivation of the agents.

---

† One might think from this that social embeddedness might be defined in terms of complexity and hence avoid the constructivist approach, but I would argue that complexity is a similar construct [13].

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 also 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 agent gets a bonus (0.2) for each 'friend' that also goes. If they stay at home they are guaranteed 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 4. The number of friendships and agents is constant across runs but the detailed structure differs. 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.

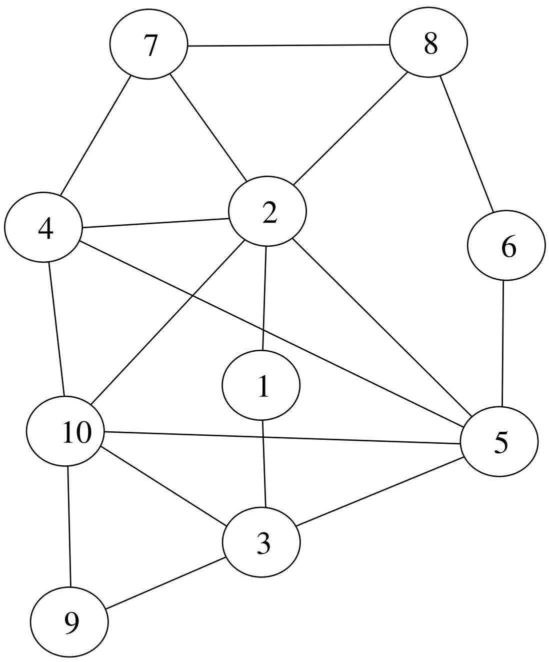

Figure 4. An imposed friendship network

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 models composed of (pairs of) expressions that represent possible behaviours in terms of what to say and what to do (its strategies). 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 population of such strategies (in this case 40). This population is very small for a GP algorithm – this is deliberate, so as to limit the explorative power of each agent to a more credible level. 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 [22] algorithm with a lot of propagation and only a little cross-over. That is 80% of the population in the next week is composed of strategies that are copies of those in the last week and 20% are formed using the tree cross-over operator from pairs of parent strategies.

Talk primitives: AND, OR, NOT, plus, minus, times, divide, boundedByPopulation, lessThan, greaterThan, saidByLast, wentLastWeek, randomIntegerUpTo, numWentLag, trendOverLast, averageOverLast, previous, quote, friendOfMine, IPredictedLastWeek, randomDecision, numWentLastTime

Action primitives: AND, OR, NOT, saidBy, wentLastWeek, previous, friendOfMine,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 5. The primitives allowed in the talk and action expressions

The formal language of these expressions is quite expressive. The primitives allowed are shown in figure 5. 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. A brief explanation of their effects during evaluation are listed in figure 6.

AND, OR, NOT – *classical logical operations of the same name*

plus, minus, times, divide – *arithmetic operations of the same name*

boundedByPopulation – *force the number output to be in the range from 0 to the number of agents in the population*

lessThan, greaterThan – *compares numbers*

saidBy, saidByLast – *what the agent indicated said or said last week*

wentLastWeek – *whether the agent indicated did last week*

randomIntegerUpTo, randomDecision – *randomly generates a number from 0 to the indicated maximum and a random boolean respectively*

numWentLag – *the number who went to the bar the number of weeks ago specified by the argument*

trendOverLast – *the bar attendance indicated by the trend over the past number of weeks specified by the argument*

averageOverLast – *the average attendance over the past number of weeks specified by the argument*

previous – *an operator to force the evaluation of the argument one week previously*

quote – *used in talk expressions to pass the subexpression of the argument literally rather than evaluate it first*

numWentLastTime – *the number of agents attending the bar last week*

IPredictedLastWeek, IWentLastWeek, ISaidYesterday – *what the agent itself predicted, did, or said last, respectively*

friendOfMine – *outputs a boolean dependent on whether the argument evaluates to the name of an agent who is a designated friend*

Figure 6. A brief explanation of the primitives that can be used to construct strategies

Some example expressions and their interpretations if evaluated are shown in figure 7. The primitives are typed (boolean, name or number) so that the algorithm is strictly a strongly-typed genetic program following [24].

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 7. 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 and Imitation

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 primitive 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 primitives 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 primitive is just evaluated as this value, but if it is a more complex expression (as a result of a quote primitive in the sending agents talk expression) then the whole expression will be substituted instead of the saidBy (or saidByLast) primitive and evaluated as such. The agent can use the output of its own messages by use of other primitives (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 weeks this expression can be crossed with other expressions in its population of strategies.

Runs of the model with and without 'listening' enabled are intended to contrast the effect of the communication on the embedding, and runs with and without 'imitation' to investigate the effect of sharing the pool of strategies explicitly. In all runs agents can follow each others actions by reference (i.e. through the use of wentLastWeek) – this differs from 'imitation' in that no transference of the content of the strategies takes place, only the results.

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.

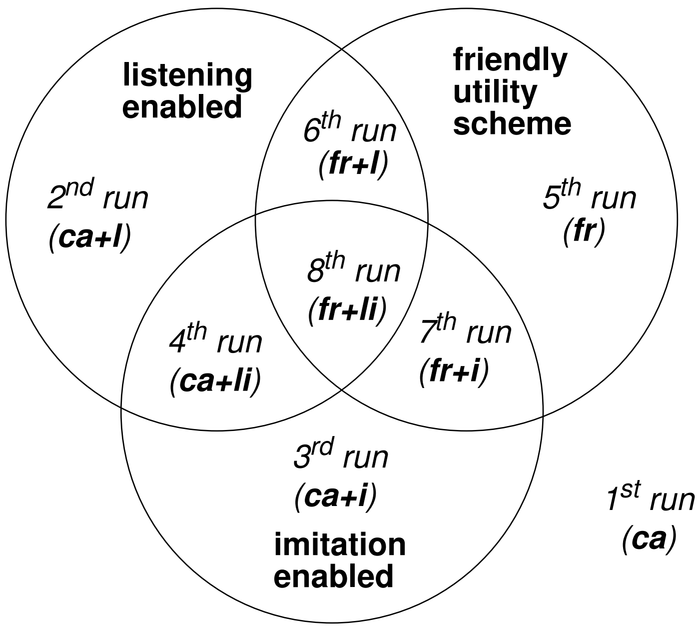

Figure 8. Variations in the 8 runs of the model

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 primitives 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 8. To ease reference to these runs I have given each a mnemonic – this is composed of two letters to indicate the utility scheme (***ca*** for 'crowd-avoiding' and ***fr*** for 'friendly'), followed by some letters to indicate whether imitation and/or listening were enabled. For example the mnemonic for the 4th run is ***ca+li***, because the crowd-avoiding utility scheme was used and both listening and imitation were enabled.

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 [27].

4.2 The Results

The complete output of the eight runs are accessible at URL: http://www.cpm.mmu.ac.uk/~bruce/socemb/data. However the reader is warned that following what is happening from these is a far from trivial matter. Bellow I summarise some of the general behaviour to provide a context for the more detailed illustrations of social embeddedness (or lack of it) that follow.

In figure 9 and figure 10 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 one's 'friends' under the friendly scheme.

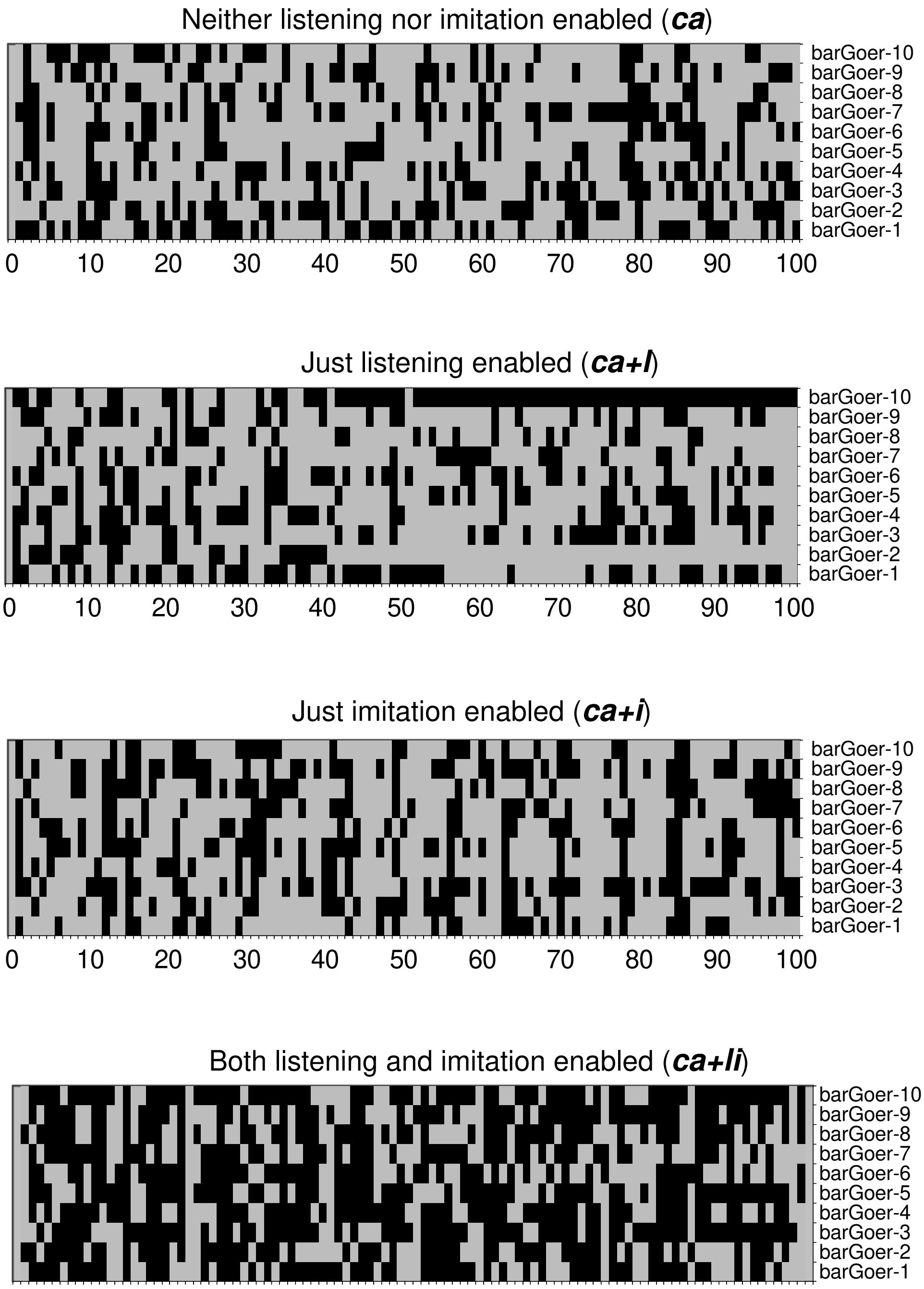

Figure 9. Attendances for the four runs under the *crowd-avoiding* scheme (grey=went, black=stayed at home)

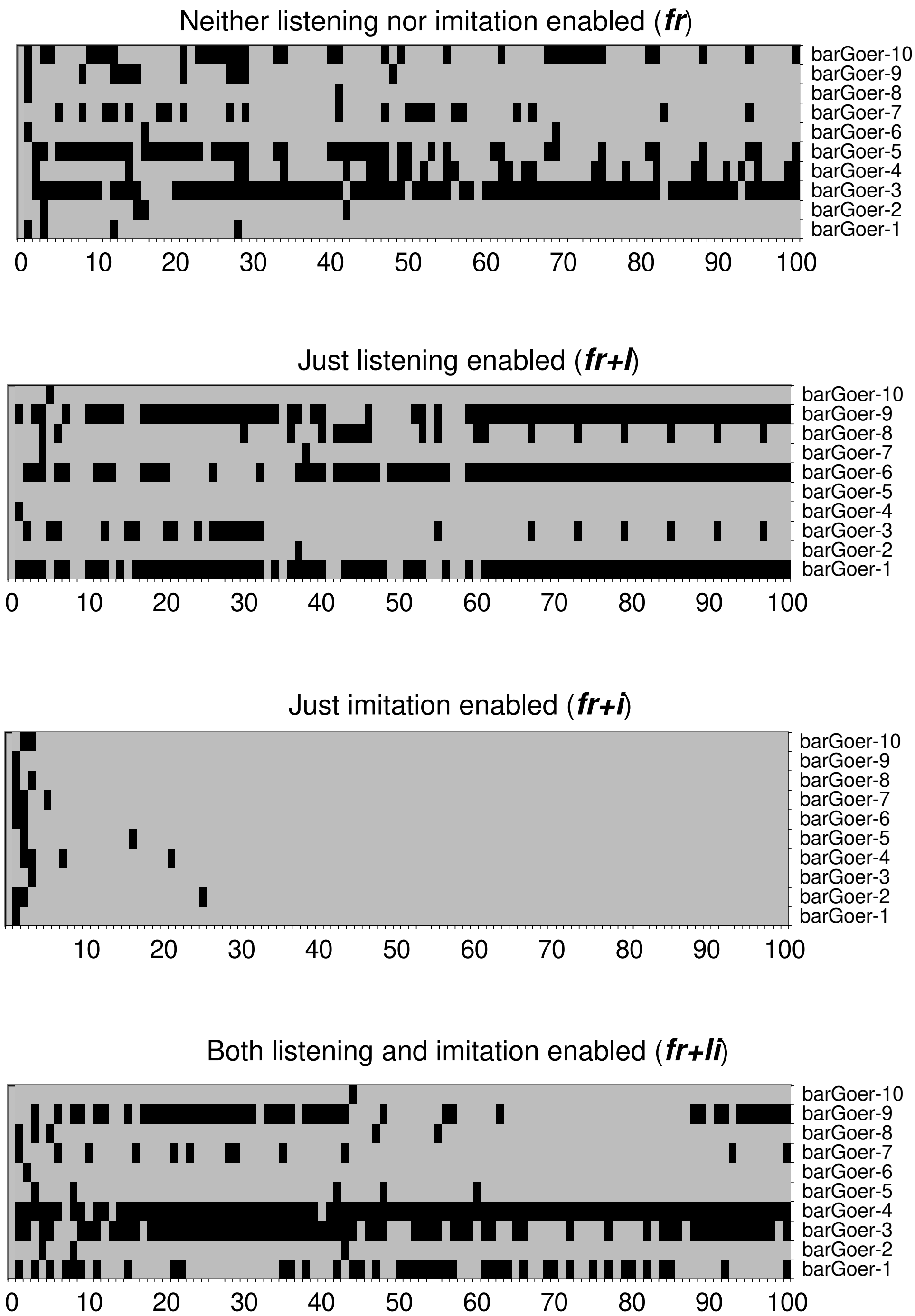

Figure 10. Attendances for the four runs under the *friendly* scheme (grey=went, black=stayed at home)

The first run exhibits the least regularity – it looks like the output from a stochastic process*. 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 the ***ca+i*** run and the uniformity of the ***fr+i*** run).

* but it may rather be the result of some sort of globally coupled chaos, as discussed in [21].

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* cannot be directly compared. Under the *crowd-avoiding* scheme (table 1) the only significant difference (at the 5% level) is between the ***ca*** and ***ca+li*** runs. In the runs using the *friendly* scheme (table 2) the only significant difference (this time at the 1% level) is between the ***fr+i*** run and the others.

ca runs	no imitation	imitation (***+i***)
listening (***+l***)	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

fr runs	no imitation	imitation (***+i***)
listening (***+l***)	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: 11, 12, 13, 14, 15, 16, 17, and 18), show some of the specific causation between the talk and action expressions of the ten agents. To keep the diagrams manageable I have limited these to the last three weeks of each run of the simulation. These figures only show the causation due to the saidBy, saidByLast and wentLastWeek primitives that are active (where by 'active' I mean a saidBy or saidByLast primitive in a simulation where listening is enabled and where it isn't logically redundant). So they do not show any causation via attendance statistics (e.g. averageOverLast, numWentLast), or the self-referential primitives (e.g. ISaidYesterday, IPredictedLastWeek and IWentLastWeek) since I wish to focus on the embedding and these are not so relevant to this concern*. In these figures there is a small box for the talk and action expression of each agent (numbered upwards from 1 to 10) – so, for example, that the topmost box under the '**A**' label represents the action strategy of barGoer-10. The numbers in the boxes are 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 an 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

* The reason for this is that the former group implement global modelling strategies and the second internal, self-referential strategies. Neither of these involve the inclusion of other specific agents' actions or utterances.

through many other agents. A detailed example of this (barGoer-6 at the end of the ***ca+l*** run) is analysed in greater detail below.

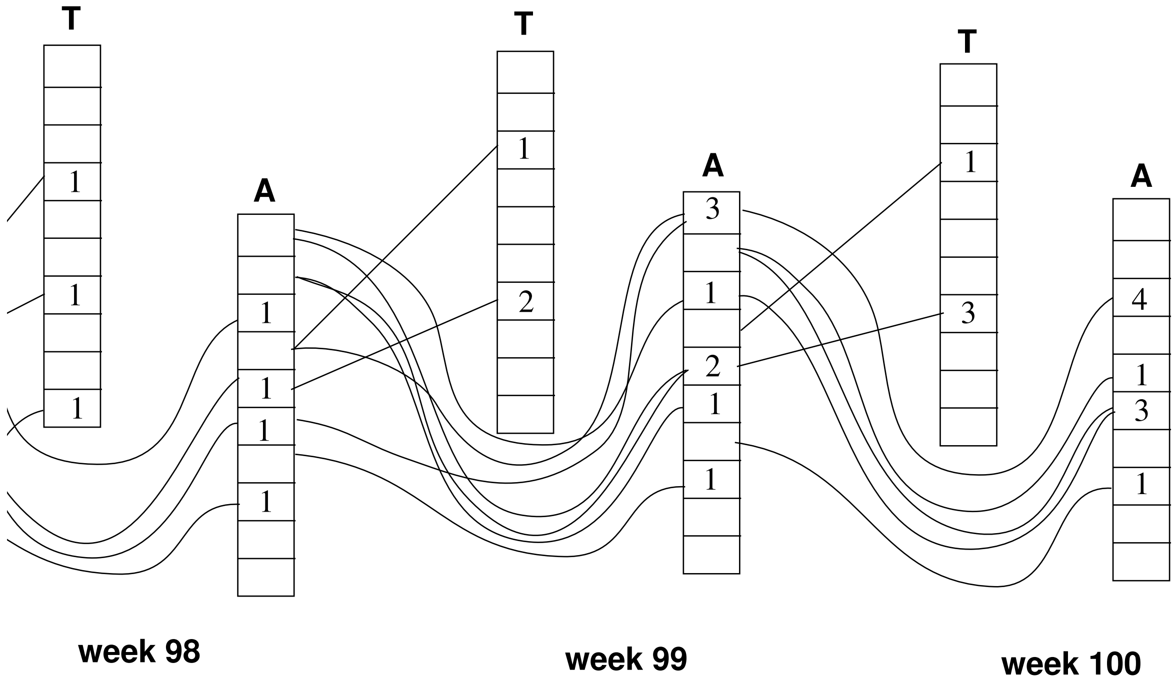

Figure 11. Causation net for run under *crowd-avoiding* scheme with neither listening nor imitation enabled (***ca***)

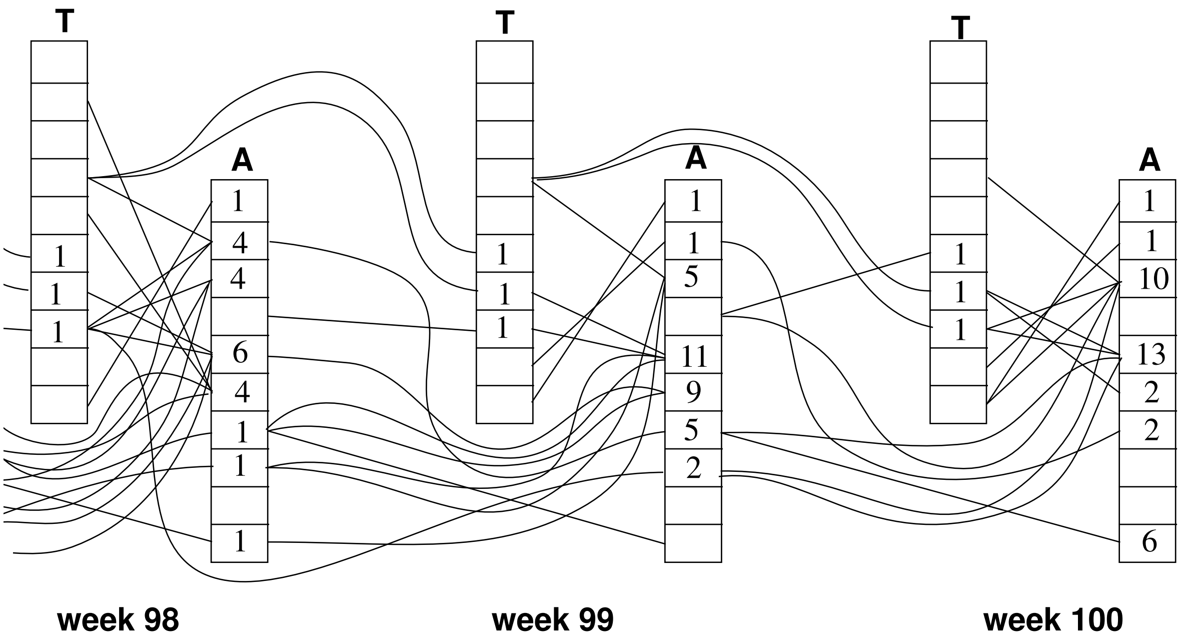

Figure 12. Causation net for run under *crowd-avoiding* scheme with only listening enabled (***ca+l***)

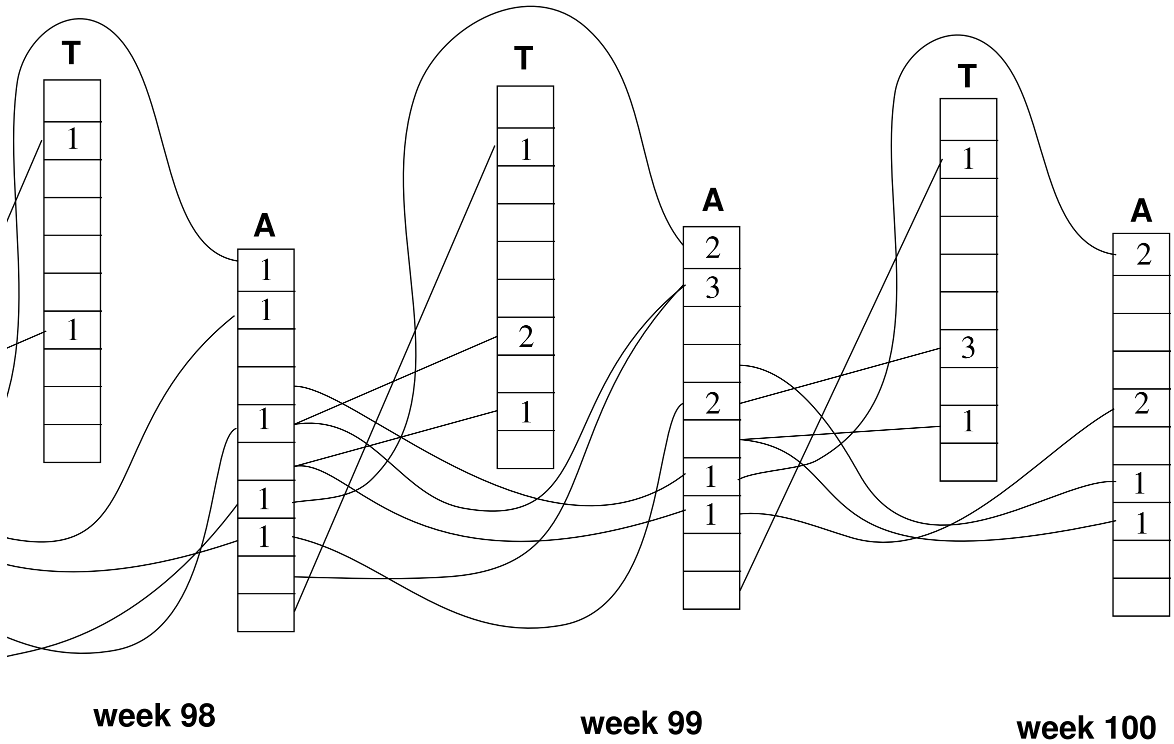

Figure 13. Causation net for run under *crowd-avoiding* scheme with only imitation enabled (***ca+i***)

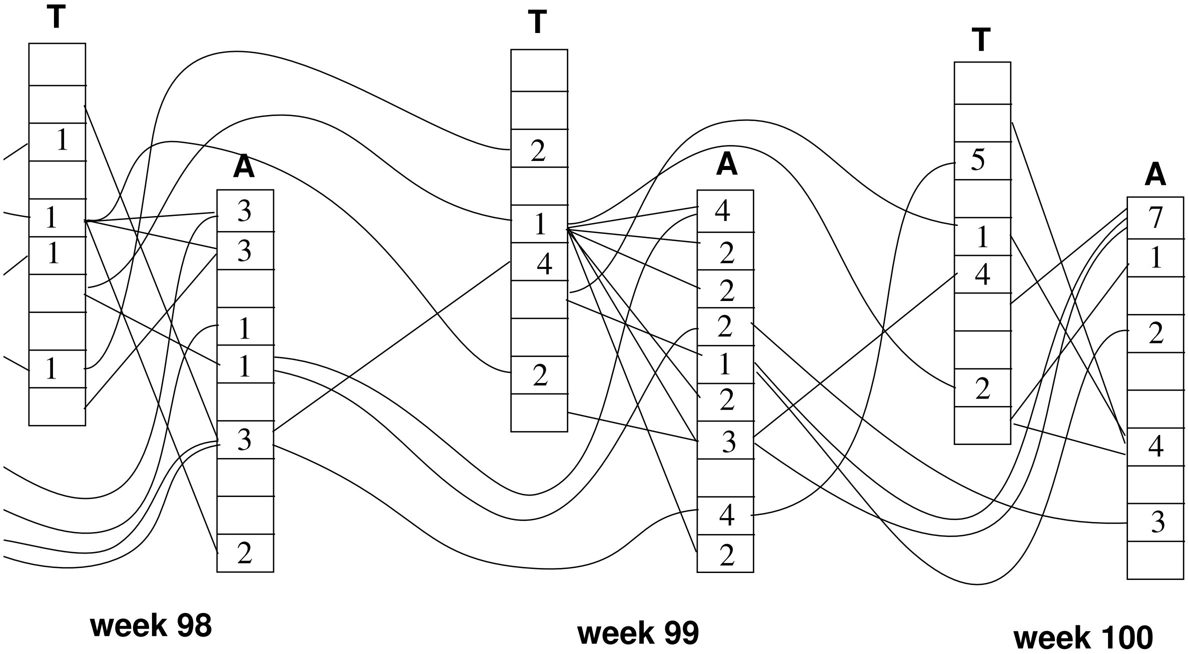

Figure 14. Causation net for run under *crowd-avoiding* scheme with both listening and imitation enabled (***ca+li***)

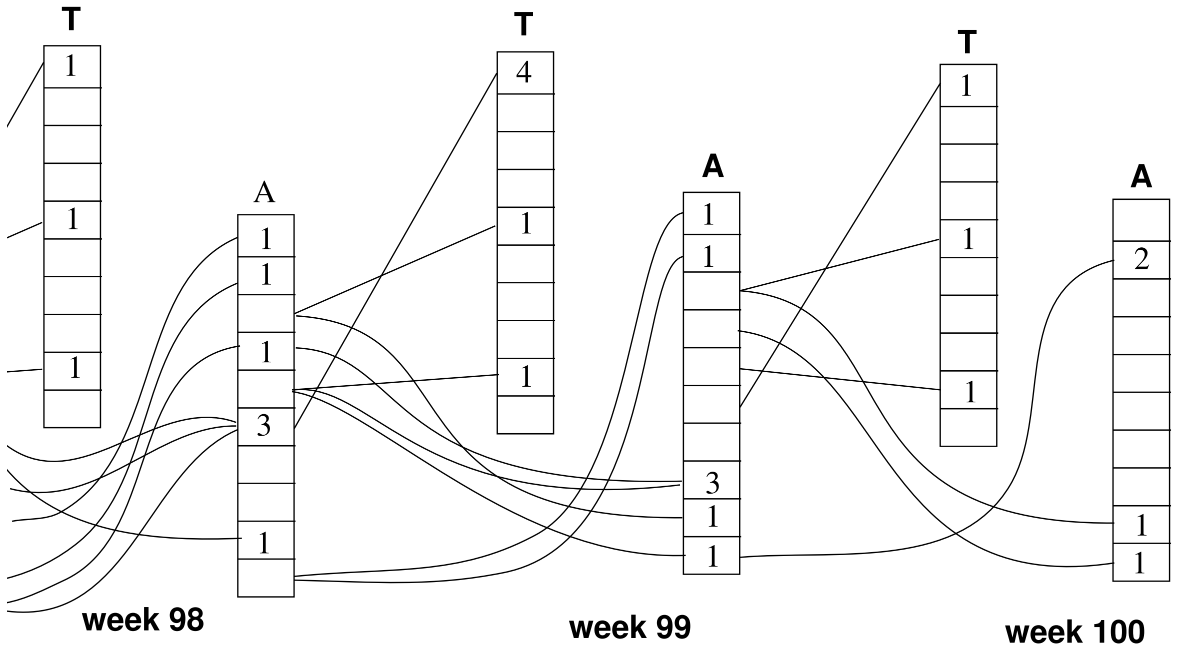

Figure 15. Causation net for run under *friendly* scheme with neither listening nor imitation enabled (***fr***)

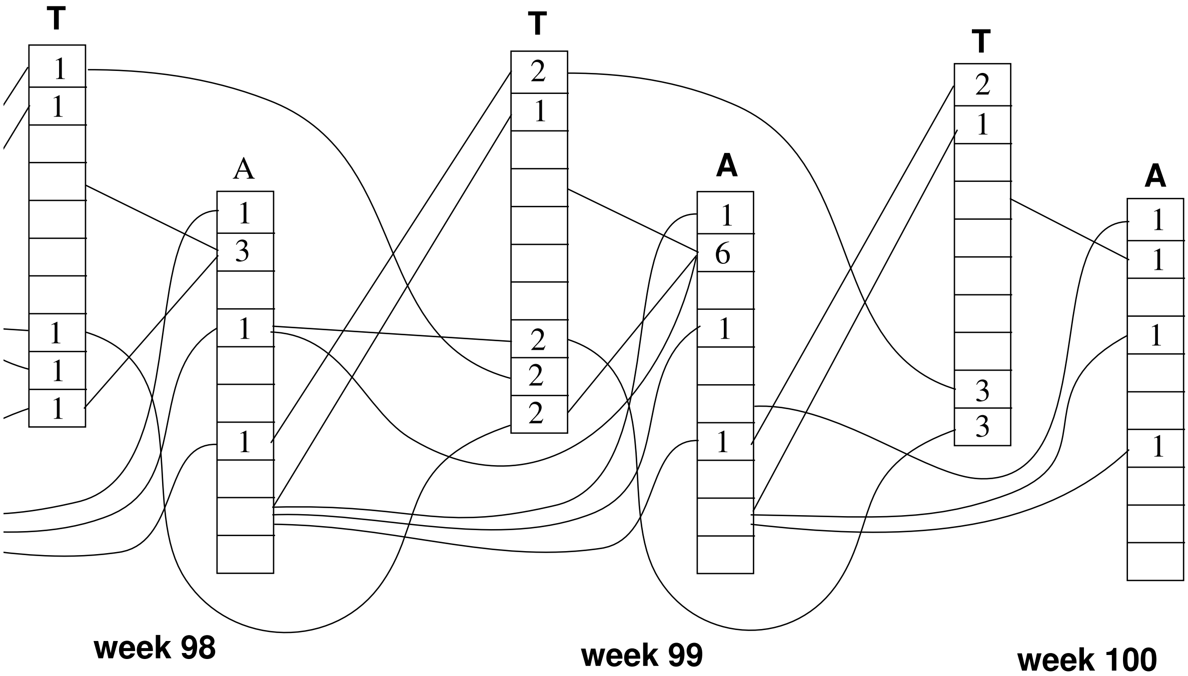

Figure 16. Causation net for run under *friendly* scheme with only listening enabled (***fr+l***)

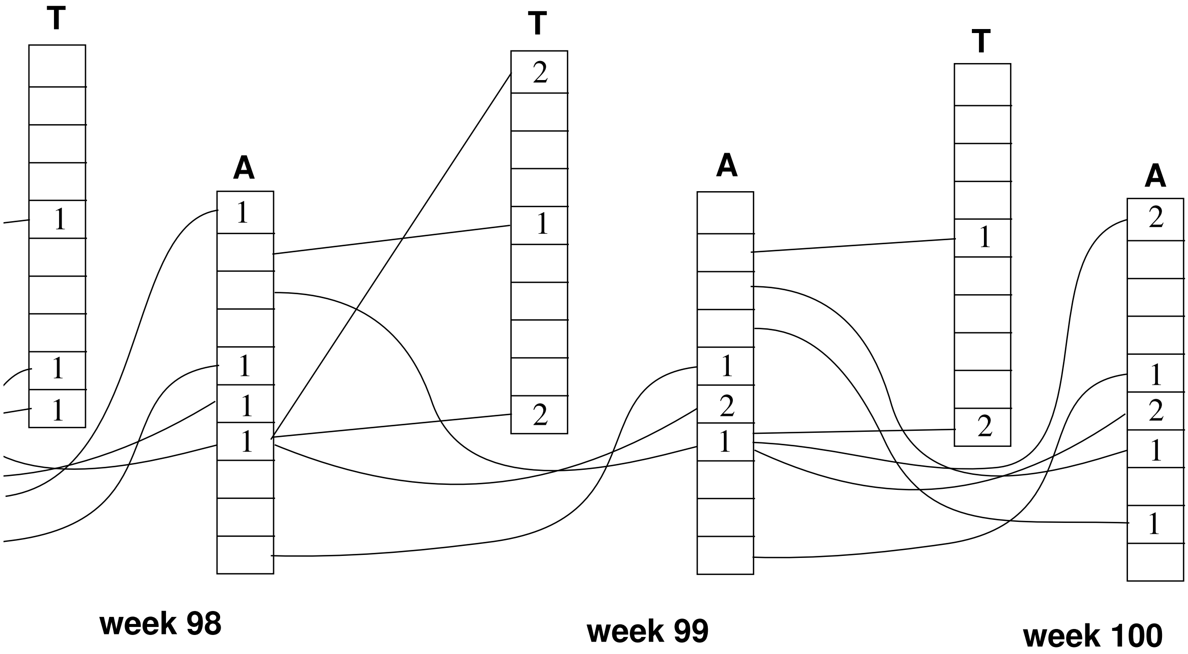

Figure 17. Causation net for run under *friendly* scheme with only imitation enabled (***fr+i***)

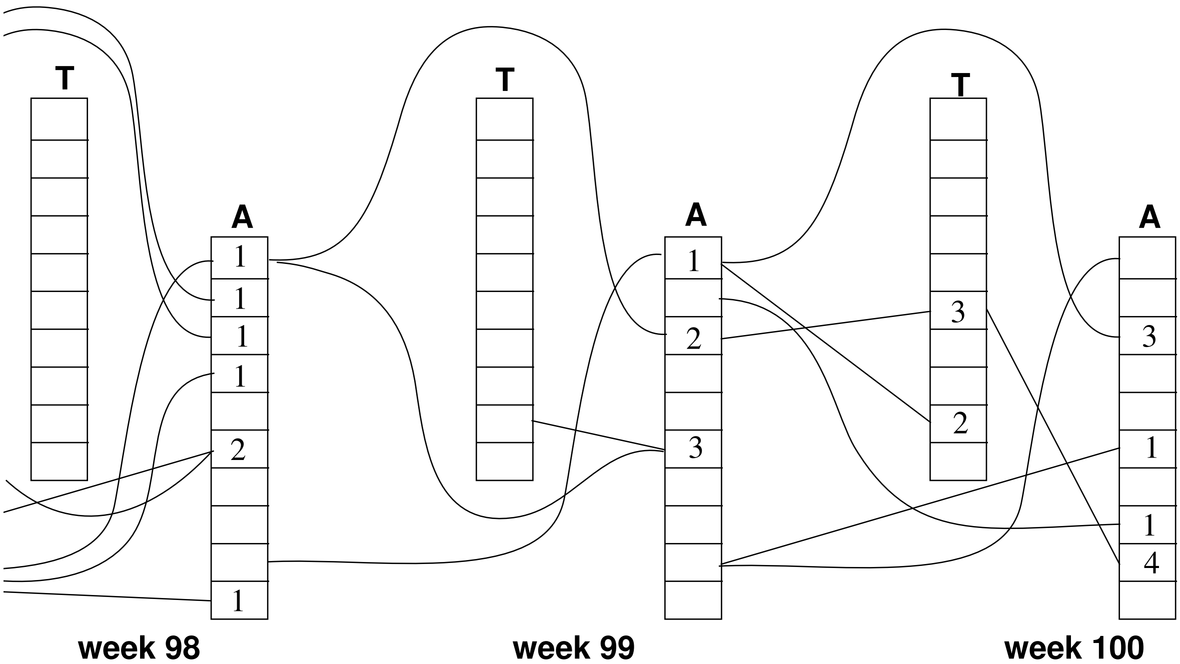

Figure 18. Causation net for run under *friendly* scheme with both listening and imitation enabled (***fr+li***)

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 are indicative only - they merely *suggest* 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. In order to check these a more detailed study of the causation involved is required.

	no imitation (***+l***)	imitation (***+li***)
crowd-avoiding (***ca***)	38.5	30
friendly (***fr***)	15.5	10

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

	no imitation	imitation (***+i***)
crowd-avoiding (***ca***)	12	12.5
friendly (***fr***)	9.5	9.5

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

4.3 More Detailed Case Studies

The above analysis of results is only suggestive as to the presence of embedding. So in order to illustrate social embedding (or the lack of it) in these simulations, I analyse a couple of detailed case studies of agent's behaviour and the causes one can attribute to it. The first example is a candidate for a socially embedded agent and the second is an example where despite the appearance of a complex web of causation there is a simple behavioural model, and hence this is not a candidate for social embedding.

4.3.1 BarGoer-6 at week 100 of the run with the *crowd-avoiding* scheme and listening only (***ca+l***)

This case is intended to illustrate the possibility of social embedding in detail. 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 bar goers 7 and 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 bar goers 2 and 6 in week 98 gives:

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

which simplifies, at last, to:

```
True
```

Even though the above trace is complex, it still ignores 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 bar goers 2 and 6 in week 98!

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

4.3.2 BarGoer-9 at the end of the run with the *friendly* scheme and listening only (*fr+l*)

In contrast to the above case-study, this example is designed to illustrate the possibility that an agent's behaviour may appear to be embedded in a complex web of social causation but that there still may be a simple explanation of its behaviour. In this case one would not say that the agent is socially embedded if one's modelling framework allowed this simpler model. Here one could say that the detailed web of causation only implemented the simpler strategy. Thus this example illustrates the importance of relativising the concept of social embedding to the modelling framework.

At week 100 the selected talk and action expressions for the 10 agents were as below (I include them for completeness, there is no need to decode these in detail).

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. I have indicated those that do not reduce to True of False, by listing them in bold. 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']]]]

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 primitives 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 depends 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 4 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 probably arises due to the fact that agents evaluate their strategies over the arises because agents evaluate their expressions only up to a horizon of five time periods into the past*.

Agent	number of friends	number of friends who are not 'loners'	attendance
barGoer-2	6	5	1
barGoer-10	5	5	1
barGoer-4	4	4	1
barGoer-5	5	3	1
barGoer-7	3	3	1
barGoer-3	4	2	1/6
barGoer-8	3	2	1/6
barGoer-6	2	2	0
barGoer-1	2	2	0
barGoer-9	2	2	0

Table 5: Number of friends, number of friends who are loners and attendance for last 30 weeks of simulation under the *friendly* scheme with listening only (***fr+l***)

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†. The friendship structure in this run was the one illustrated in figure 4. Agents barGoer-3 and 8 are 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 runs of the simulation that showed a high degree of social embeddedness exhibit most of the predicted effects which were listed (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

* It is noticeable that in the earlier part of this run these two agents had broadly complementary patterns of attendance, which is understandable due to the friendship structure.

† Contributions towards the Society for the Abolition of Agent Depression should be sent via the author.

intuitions on the topic; my ideas about social embeddedness and the model have themselves co-developed. In particular:

- the expressions that the agents develop strategies that are opportunistic – they do not reflect their social reality but rather constitute it as causal elements;
- the strategies 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 frequently do use incorporate information about the communication and actions of other individual agents instead of attempting to predict their environment using global models – 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 19. Here we can see that the primitives for concerning others actions and utterances are heavily selected for, while those involving global statistics, random elements or backward looking primitives are selected against;
- 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];

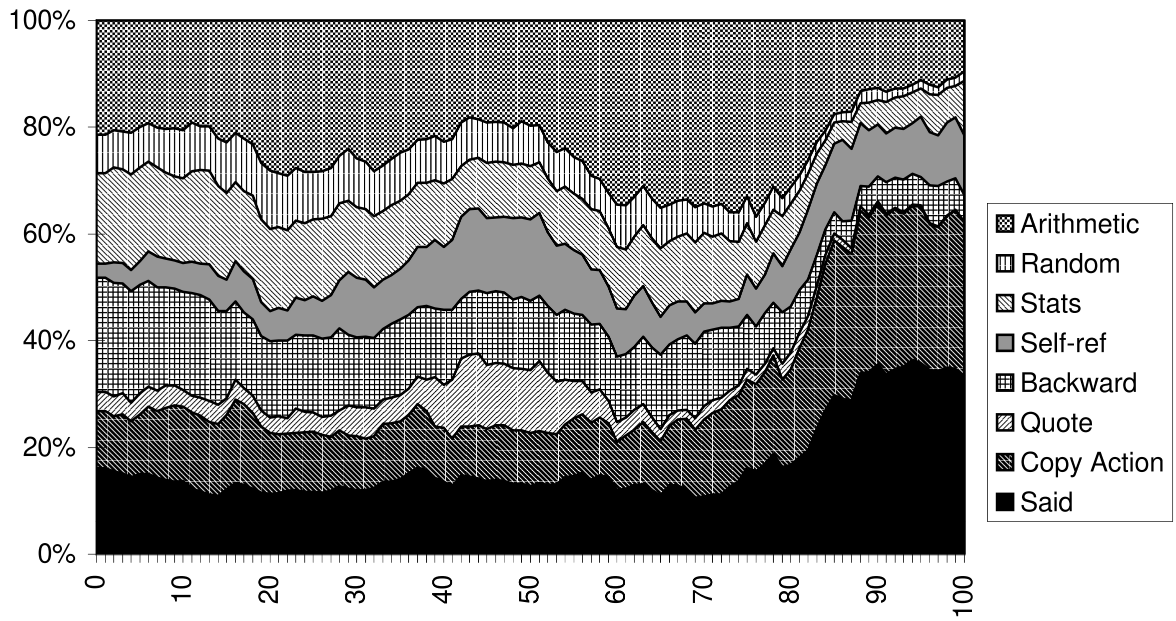

Figure 19. Distribution of the relative proportions of some primitive types in the run using the *crowd-avoiding* scheme with only listening enabled (***ca+l***)

It is 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 effect 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 maybe this evolution would have taken advantage of the behaviour of the other agents; this would be analogous to the way different mechanisms in one organism develop so that they have multiple and overlapping functions that defy their strict separation [33];
- 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 modelling by agents, i.e. that there is no practical limit to the variety or complexity of such models – if the space of possible models was essentially small (so that an approximation to a global search could be performed), then the optimal model of the society that the agent inhabited would be feasible for it;
- mechanisms for social distinction (e.g. a naming mechanism) and hence the ability to develop the *selective modelling* 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), so that individual intelligence might both have enabled the development of social embedding as well as being selected for it (as in the ‘social intelligence hypothesis’ discussed in [23]).

What is unclear from the above model and analysis is the role that imitation plays in the development (or suppression) of social embeddedness, particularly where both imitation and conversational communication are present. In [9], Kerstin Dautenhahn 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 that, at least sometimes, imitation may have a role in *simplifying* social situations so that such embedding does not occur.

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.

Its application may have the most immediate impact upon our modelling methodology. 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 *suppressing* 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.

* This is similar to a remark in [1].

Acknowledgements

Thanks to Scott Moss and 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

(All CPM Reports can be accessed from the CPM web site at: www.cpm.mmu.ac.uk)

[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] Clancy, W. J. (1997). *Situated Cognition*. Cambridge University Press, Cambridge.

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

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

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

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

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

[14] Edmonds, B. (1998). Gossip, Sexual Recombination and the El Farol Bar: modelling the emergence of heterogeneity. Computation in Economics, Finance and Engineering: Economic Systems conference (CEFES'98), June/July 1998, Cambridge, UK. Also available as CPM Report 97-31, MMU, Manchester, UK.

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

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

[19] 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.

[20] Harnad, S. (1990). The Symbol Grounding Problem. *Physica D*, 42: 335-346.

[21] Kaneko, K. (1990). Globally Coupled Chaos Violates the Law of Large Numbers but not the Central Limit Theorem. *Physics Review Letters*, 65:1391-1394

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

[23] 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.

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

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

[26] 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.

[27] 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

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

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

[30] Suchman, L.A. (1987). *Plans and Situated Actions*. Cambridge: Cambridge University Press.

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

[32] Vera, A. H. and Simon, H.A. (1993). Situated Action: A Symbolic Interpretation. *Cognitive Science*, 17:7-48.

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

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