Dynamics of Shared Mental Representation: What can a simple network of agents tell us?

Lynette Shaw†, Sarah Wise‡, Micael Ehn§, Ingrid van Putten‖

Abstract

This paper develops a simple model of how individuals automatically and unconsciously use social information feedbacks from other individuals in order to determine the mental representation they will impose upon a social situation. An agent-based modelling approach is used to demonstrate how these learning processes, when carried out in an inter-subjective context, are sufficient to generate a number of dynamics that characterize real social systems. Results indicate that both network structure and updating strategies significantly determine the pattern of mental representation adoption across the set of agents. Significant findings include the non-trivialness of reaching full consensus in a group, the emergence of distinct sub-groups and cultural “brokers” between them, and the variable ability of a single agent acting independently of social feedbacks to drive the entire system toward consensus.

Introduction

Much research has focused on getting better insight into how we develop an understanding of the world. After all, the world is not a self-evident place. All information we receive from the environment requires us to make inferences about what’s going on and ascribe meaning to what has happened. It is on these inferences and ascriptions of meaning that we base all our subsequent opinions and actions.

There is a firmly established body of research in human cognition and the so-called social cognitive neurosciences (Lieberman 2007) that indicates these interpretations of “what is happening” arise from a set of unseen background assumptions we automatically and unconsciously impose on the situations in which we find ourselves.

A large amount of evidence amassed from theoretical and experimental work (e.g. Dijksterhuis, Chartrand, & Aarts, 2007; Ferguson, 2007) indicates that much of human cognition and action arises out of automatic, unconsciously executed

† Department of Sociology, University of Washington
‡ The Department of Computational Social Science, George Mason University
§ The Centre for the Study of Cultural Evolution, Stockholm University and Mälardalen University
‖ Commonwealth Science and Industry Research Organisation, Marine and Atmospheric Research, Hobart, Tasmanian, Australia
processes. In social psychology, discussions of these processes, which have been described in terms such as “efficient,” “fast,” and “unconscious” (Moors & Houwer, 2007), have fallen under the heading of the “automaticity” (Bargh, 1982).

It is through these automatic processes that we impose mental representations, or frameworks of meaning, onto our experiences. It is these representations which subsequently shape the values, opinions, attitudes, and behaviours of which we can become consciously aware. In the history of social science, these representations have been referred to in such terms as categories (Allport, 1954), schemas (Bartlett, 1932; Taylor & Crocker, 1981), cognitive scripts (Abelson, 1981), theories (Murphy & Medin, 1985), conceptual metaphors (Lakoff & Johnson, 1980), and subjective mental models (Holland, et. al. 1986). The reality of this “dual-process” nature of human cognition has recently begun to be carried into various areas of social science including sociology (Vaisey, 2009) and behavioural economics (Kahneman, 2003).

Throughout the social sciences, very similar conceptualizations of this inherent feature of human cognition have also been invoked toward a striking diversity of explanatory ends. Examples include North’s mental models (North, 1994), the ‘frames’ of social movements’ studies (Snow, Rochford Jr., Worden, & Benford, 1986), and even the ‘choice-sets’ that arise under bounded rationality (Kahneman, 2003). Similar concepts have also been applied to areas of culture (Dimaggio, 1997), and organizations (Meyer & Rowan, 1977). An exceptionally comprehensive paper by Howard (1994) also makes a powerful argument for the incorporation of mental representations and related processes into explanations of social structures.

Even though much has been written on mental representations, automaticity, and the social construction of reality, considerably less attention has been paid to how frameworks of meaning come to be shared across a group. In this research we develop a model of how individual humans automatically and unconsciously use social information feedbacks to settle upon the mental representation they impose upon a social situation. We then forge the often sought link between the micro and macro level (Coleman, 1990) by investigating the complex social level dynamics that emerge as a result of this simple, individual-level process. Toward this end, we use an agent-based modelling approach to demonstrate how this model of individual cognition can lead to interesting and well supported statements about how the social world operates.
Methodology

The foundational model, though firmly grounded in established research on human cognition, is very simple. It moves away from more complicated models of the social world and instead seeks to understand how much of social system dynamics can arise from the relatively straightforward process of human learning in an inter-subjective context. With very few assumptions, this model is able to parsimoniously generate and explain previously simulated (e.g. Axelrod 1997; Boyd and Richerson 1985; Henrich and Boyd 1998) and observed social phenomenon. The basic conceptual framework is shown in Figure 1.

![Confirmation of Mental Representation Imposition](image)

**Figure 1:** Conceptual model of mental representation imposition using inference validation.

The model shown is similar to Bruner’s (Bruner, 1957) model of how humans categorize their perceptions. Our model is, in fact, a very general model of learning. Though the dynamics of learning is a very well-studied subject, the special condition of human cognition in a social context is a critical variation on established work. Most often, learning is conceptualized as occurring against the backdrop of a fixed environment, where information feedback is relatively invariant. However, often behaviours exist that clearly adapt in different degrees to the changes in given circumstances.
In the social case the effect of actor adaptation is in two directions. Not only do actors adapt to their own environment, they are also part of the environment to which other actors adapt. In order for an actor to make valid inferences about the potential action of surrounding actors, an actor needs to be able to assume the mental models of those surrounding actors. Simultaneously, all those that surround the actor need to do the same. The inherently inter-subjective nature of this social circumstance has long been identified as a key insight to understanding the social world (classic examples include Berger & Luckmann, 1966, Goffman, 1959, and Habermas, 1984).

In this current study we use this understanding of the social world to further explore the effect of actors, who are each characterised by a set of mental representations, on the dissipation of mental representations in social systems characterized by different social network structures. We explore the dynamics of five different social network configurations using this agent-based approach. We have called the agents in the simulations SocialActors.

As mentioned above, the SocialActors are each characterised by a set of some established number of mental representations that can characterize the situation in which they find themselves. The mental representations are initially assigned a random weighting and the SocialActors are placed in a social setting amongst a number of neighbouring SocialActors who each have their own set of weighted mental representations.

In one set of simulations, every SocialActor is initialized to play a ‘majority strategy’ and in the other set they are initialized to play a ‘probabilistic strategy’. There are two key processes that differ between the two strategies: how the SocialActors update their mental representations and what mental representation the SocialActors choose to play themselves. The time step on which the simulation operates is an abstract one with a ‘step’ in the simulation representing an interaction between every SocialActors and its neighbours. These different processes that characterise the two strategies play a central role in the simulations.

Under the majority strategy, SocialActors play deterministically. They are initialized to have “observed” some number of instances of each representation being played by their neighbours in the past. Every turn, SocialActors observe the mental representation of each of their connected neighbours. They subsequently update their
own weighting of the mental representation being played by their neighbours by one for every positive observation of this mental representation with their neighbours. Each SocialActor then plays its highest-weighted mental representation.

For the probabilistic strategy, SocialActors choose which representation to play probabilistically. They are initialized to weight each representation randomly between 0 and 1. Every time they observe another of the connected SocialActors play a certain mental representation, they update their own weighting of this mental representation, increasing the old weight of that representation by an activation factor. Given an activation factor \( a \), a maximum probability \( p \), and a played framework with weight \( w \), an agent will set the new weight of that mental representation to:

\[
\text{new weight} = \min(p, w \times a)
\]

In the probabilistic strategy, when an agent observes a certain mental representation, its weighting for all other mental representation will decay. With a decay factor \( d \), a minimum probability \( p \), and an unplayed mental representation \( a \), the SocialActor will set the weight \( (w) \) for the mental representation to:

\[
\text{new weight} = \max(p, w \times d)
\]

The imposition of a minimum possible weight for representations indicates that no mental representation is ever regarded as completely impossible, merely extremely improbable. When a SocialActor selects the mental representation it plays, it will select a given mental representation with probability \( w/s \) where \( s \) is the sum of all of the weights of different mental representations of that SocialActor.

As mentioned before, the SocialActors are organized into different social networks representing different social circumstances. The five different networks include a simple ring network, a Small World network (Watts and Strogatz 1998), an Erdős–Rényi network (Erdős and Rényi 1959), a fully connected network, and what we call the Broadcaster network. A ring network is just a network wherein SocialActors are situated in a ring and only connected to their two nearest neighbours. A Small World network consists of a ring of connected SocialActors with connections between a few
non-neighbouring actors which dramatically decrease the average path length between any two actors. An Erdős–Rényi graph is constructed by randomly linking pairs of nodes in a graph. A fully connected network is one in which every node is connected to every other node. The Broadcaster network developed here features a Small World network with a single highly connected agent who always weighs all but one framework at zero, independent of social feedbacks. This type of network is meant to simulate a number of real world social circumstances where there is, for instance, a highly visible leader, a charismatic individual, or simply an individual with very fixed ideas. While the first four networks are what may be termed ‘conventional’, the broadcaster network is a variation on the network structure that does not appear in the social modelling literature and fits an original and interesting type of social circumstances.

Results

Our preliminary results reflected the importance of both the social network and the SocialActor strategy employed in the simulation. The ultimate emergence of distinct social groups showed that the simple behaviour of the agents produced structures that were not built into the simulation. This is a hallmark of complexity and supports our theory that the social systems we observe in our own world can be explained by relatively simple rules similar to the ones presented here. Some initial data characterizing the different systems considered here are provided below in Table 1.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Ring</th>
<th>Small world</th>
<th>Fully connected</th>
<th>Erdős–Rényi</th>
<th>Broadcaster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of edges</td>
<td>100</td>
<td>140</td>
<td>4950</td>
<td>200</td>
<td>146</td>
</tr>
<tr>
<td>Average pathlength</td>
<td>12.7</td>
<td>2.5</td>
<td>1</td>
<td>1.8</td>
<td>2.4</td>
</tr>
<tr>
<td>Clustering coefficient</td>
<td>0.000</td>
<td>.005</td>
<td>1.000</td>
<td>.016</td>
<td>.011</td>
</tr>
<tr>
<td>Average number of representations present after convergence</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1: All statistics are averages from 10 runs performed on groups of 100 SocialActors. Networks were unchanged throughout the run, with no edges being added or deleted. SocialActors had 3 available mental representations.
Across the different network types, distinct patterns of representation adoption occurred under both Majority and Probabilistic strategies. Overall, the most significant result this data suggests is that the normalizing force agents exert upon one another is strongly dependent upon the connectedness of the graph within which they are situated. In general, the more connected a graph is, the more likely agents are to be part of a single large group of like-minded individuals. Conversely, networks with longer pathlengths and lower clustering coefficients often resulted in stable subgroups of individuals with similar mental representations. Even though all agents strove to match those in their local neighbourhood, coordination across the entire network proved to be a nontrivial issue.

At present, the average number of representations acts as a very rough indicator of the “grouping” behaviour that emerged for different network structures and agent strategies. Future refinements of this research will likely involve developing more fine-grained measures which better capture the existence of distinct subgroups at time of convergence. When the above quantitative results are considered in conjunction with qualitative evaluations of system behaviours along different networks, however, the grouping trends described earlier become quite evident.

In both ring and small world networks, cohesive communities of agents emerged regardless of the selection strategy employed by the agents. In both cases, stable groups of individuals with similar mental representations emerged just by virtue of the learning that took place along the given network structure. Illustrations of this typical behaviour are provided in Figures 2a and 2b.
As to be expected, the initial subgroup formation patterns of Small World networks held for Erdős–Rényi graphs. However, the random rewiring of the network produced a flow of learning that resulted in the initial morphology of the community being quite important. Sometimes a representation that took hold of some cohesive group of
agents near the beginning of the simulation propagated out from this base of support, eventually eliminating a rival representation.

The more connected a network became, the more significant the difference in the majority and probabilistic strategies were. In the case of probabilistic agents, fully connected networks produced “consensus” or the dominance of one mental representation across the entire community. Under the same conditions in the case of majority strategy, however, all representations continued to persist in small groups of agents. Figures 3a and 3b illustrate a highly, though not fully, connected system of probabilistic actors reaching consensus across the entire group.

Figure 3a: Initial, randomly assigned state of a highly connected network of Probabilistic actors
Figure 3b: Same network stabilizing at a full “consensus”

In broadcaster networks, broadcasters (agents playing one representation permanently, independent of social feedbacks) were situated in moderately connected Small World networks, a structure which was established as usually producing stable patterns of subgroups. Early results indicate that in some cases where agents are playing the probabilistic strategy, a broadcaster connected to a minority of the population is able to drive the system toward consensus or near consensus. Figures 4a and 4b illustrate how, in a manner that is likely very contingent upon initial conditions, a single broadcaster agent connected to only 15% of the population can strongly drive it toward consensus on the representation it itself is playing.
Figure 4a: Initial state of a moderately connected Small World network of Probablistic actors wherein 15% of the population is connected to a “broadcaster” agent playing red

Figure 4b: Same system approaching convergence upon a final state. At a later point, this system reaches full “consensus” upon red
In addition to expected differences in the ability of systems to reach consensus, another interesting and unexpected phenomena that emerged was that of “brokers”\(^1\) – agents situated at the boundaries between groups who stabilized into a pattern of perpetually “switching” between the representations of the groups to which it was connected. This behaviour emerged when agents sustained equal or nearly equally weighting for two or more representations due to their positioning in between agents who had unambiguously converged upon different representations. The characteristic “switching” behaviour of brokers was only observable in the probabilistic strategy as agents in the majority strategy with equal or near equal weighting of frameworks always played the framework with the highest weight, no matter how close the weights were, and always broke ties in favour of the representation they had previously played.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Ring</th>
<th>Small world</th>
<th>Fully connected</th>
<th>Erdős–Rényi</th>
<th>Broadcaster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of edges</td>
<td>100</td>
<td>140</td>
<td>4950</td>
<td>200</td>
<td>146</td>
</tr>
<tr>
<td>Average pathlength</td>
<td>12.7</td>
<td>2.5</td>
<td>1</td>
<td>1.8</td>
<td>2.4</td>
</tr>
<tr>
<td>Clustering coefficient</td>
<td>0.000</td>
<td>.005</td>
<td>1.000</td>
<td>.016</td>
<td>.011</td>
</tr>
<tr>
<td>Average number of ‘brokers’ in final configuration</td>
<td>24.4</td>
<td>n/a</td>
<td>15.1</td>
<td>n/a</td>
<td>10.6</td>
</tr>
</tbody>
</table>

**Table 2** All statistics are averages from 10 runs performed on groups of 100 SocialActors. Networks were unchanged throughout the run, with no edges being added or deleted. SocialActors had 3 available mental representations.

As seen in Table 2, ring networks produced considerably more brokers than small world networks did, following an overall trend that showed longer path lengths and higher clustering coefficients being associated with fewer brokers. Given that these network characteristics are also indicative of fewer subgroups, this trend is not surprising given that the presence of fewer subgroups entails less boundaries at which brokers can emerge. The fact that these brokers frequently but not necessarily arose

\(^1\) Although ‘broker’ is a term used in network analysis to indicate nodes with high ‘betweenness’ values, in this current study it simply refers to the agents who mediate between groups. More specifically it refers to the “switching” behaviour typical of the broker agents in some of our models.
at the borders between groups also points to the importance not just of position but also individual history of interactions in the generation of this sort of behaviour.

A final measure considered captures the degree to which an agent “agrees” with its neighbours, which we here call “contentedness” after the tradition of Schelling. This number was simply the proportion of an agent’s neighbours that adopted the same representation as the agent in question, with higher numbers indicating greater contentment.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Ring</th>
<th>Small world</th>
<th>Fully connected</th>
<th>Erdős–Rényi</th>
<th>Broadcaster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of edges</td>
<td>100</td>
<td>140</td>
<td>4950</td>
<td>200</td>
<td>146</td>
</tr>
<tr>
<td>Average pathlength</td>
<td>12.7</td>
<td>2.5</td>
<td>1</td>
<td>1.8</td>
<td>2.4</td>
</tr>
<tr>
<td>Clustering coefficient</td>
<td>0.000</td>
<td>.005</td>
<td>1.000</td>
<td>.016</td>
<td>.011</td>
</tr>
<tr>
<td>Average contentment</td>
<td>.830</td>
<td>.470</td>
<td>.820</td>
<td>.488</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3 All statistics are averages from 10 runs performed on groups of 100 SocialActors. Networks were unchanged throughout the run, with no edges being added or deleted. SocialActors had 3 available mental representations.

The choice of agent strategy also had an important impact on the number of representations on the average contentedness of agents. Probabilistic updating allowed for unpopular representations to be driven out of the system, while majority updating preserved them. Across the board, probabilistic updating resulting in reasonably contented agents while majority updating resulted in highly discontented agents. Additionally, in the fully connected network, probabilistic agents were able to reach consensus and as a result, perfect contentment across the system was achieved. Given that majority agents were not able to reach consensus under the same conditions, such contentment levels could not be achieved, and interestingly, they were ultimately less content than they would have been in a less connected network structure.

**Discussion**

In this study we build on existing research in human cognition to create an analytically tractable version of a social actor that has been alluded to by both classical theorists and modern empirical social researchers. We commenced with a
parsimonious and highly generalizable set of assumptions about individuals in order to develop a set of statements about social dynamics that might be applied at many levels of social interaction and in many different substantive contexts. Toward this end, an agent based simulation of the development of mental representations in a social context was developed. We showed that the convergence of mental representations in a social setting is a function of both network configuration and the strategy by which actors in the network are conceived of imposing their own mental representations.

Though our model assumptions are extremely simple, our results indicate many interesting phenomena which resemble well-established results in social simulation. For example, findings concerning the development of subgroups are similar to those discussed in Axelrod’s (1997) “Dissemination of Culture”. Axelrod considers cultural dissemination using evolving “belief vectors” and defines two actors as belonging to the same cultural subgroup if they have identical belief vectors. Axelrod also assumes the presence of homophily, that the probability of two actors interacting is positively correlated with the similarity of their belief vectors, and concludes that this element is important to explaining the formation of subgroups. We are able to generate the same pattern of subgroup formation through learning dynamics and network structure alone without any assumptions of homophily. Furthermore, this result is accomplished with a much simpler model of the mental representations underlying beliefs that is nevertheless much more concordant with research demonstrating that beliefs do not exist in isolation but as part of a coherent system (Lakoff & Johnson, 1980).

The emergence of brokers is also reminiscent of some of Epstein’s findings in “Learning to Be Thoughtless” (Epstein, 2001). In his model, Epstein seeks to get at how norms spread across populations of agents who behave like “lazy statisticians” in their attempts to match how others around them are behaving. In addition to also finding the development of subgroups, Epstein finds buffers of agents with higher uncertainty in their behaviour that form around agents who have settled in upon the locally dominant model. Our model also produces such agents in the form of the so-called “brokers,” but does so with less complicated updating rules than the one’s Epstein uses involving varying search radiuses.
It is of further interest that this model is able to conceptually bridge these two separate models of norms and culture while also pointing to still further possible links with well-known social simulations. One brief example of other such potential connections is to the inductively reasoning agents, which were designed in part based on research into mental models, that Arthur considers in his famous “El Farol Bar Problem” (Arthur, 1994). It is expected that more such bridges between well-known simulations will be found as this research progresses.

In terms of the continuing development of this agent based model, there are many potential directions to consider. For instance, modelling different authority structures would be quite straightforward via the incorporation of unidirectional networks. The effect of removing or adding different types of links on the spread of representations in a network could get at the continuously changing patterns of social interactions among friends and enemies. The strategic removal of highly central nodes would very likely be able approximate the effects of removing a leader from a group. Including varying weights to links might also be a way to consider the influence of strong emotion or levels of trust in an interaction. Beyond network structure, other aspects of human cognition, such as confirmation bias or “gatekeeping” (Bruner, 1957) and the nested nature of frameworks (Taylor & Crocker, 1981), represent potential refinements to the model. The incorporation of non-social sources of feedback would also likely result in an entirely new arena of learning dynamics to investigate.

The ultimate goal of most any social simulation endeavour, however, is real world relevance. Fortunately, within the wider context of empirical research our simulations have many potential areas of application. The scenarios modelled here could, for instance, represent a small group of strangers settling upon a consistent way of interacting with one another, a team in an organization trying to figure out how to approach a problem, social movement members trying to sway the larger collective toward a certain framing of an issue, or a society which is in the process of establishing a stable culture. In all of these different cases, the simulations presented in this study offer many hypotheses about how such dynamics are likely to play out given a set of structural constraints. On the whole, there exists an exceptionally wide and varied set of empirical cases against which these models could be validated.
Moving further out into the realm of policy development, as these models become more sophisticated and are successfully vetted against empirical data, there are many potential ways they might be used by policymakers. One example could be predicting the success of a public awareness campaign given certain known characteristics of a community structure. Other examples could include anticipating patterns of cultural assimilation among immigrants, modelling the potential spread of a radical social movement, or simulating the situations under which a major environmental or economic shock would lead to widespread social disorder.

In summary, this work represents the initial development of a very simple but general conceptual model of the social world. Already it has demonstrated how the network of who interacts with whom and the manner in which individuals adapt their worldviews to new information might be sufficient to explain whether or not a group shares an understanding of the situation they face or is instead marked by fundamentally different conceptions of what is happening. The results developed here have potential relevance not only to many different bodies of existing social simulation work, but also to many areas of contemporary social theory and research. The ultimate goal of this work is to open up new lines of dialogue across the substantive and conceptual divides that presently mark much of the social sciences. These early results provide a very encouraging indication that it may eventually be able to do just that.
References


