18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Modeling multidirectional, dynamic social influences in social networks

Van Rooy, D.

Department of Psychology, Faculty College of Medicine, Biology and Environment, The Australian National University, Australian Capital Territory
Email: dirk.vanrooy@anu.edu.au

Abstract

The study of social influence seems to have developed along two parallel, but largely independent lines of research. On the one hand, research in sociology and physics has focused on the macro-level, by studying dynamics of opinion flow within extended social influence networks and using aggregate-level variables (i.e., the proportion of a population in a particular state), with little regard for individual psychological processes working at the micro-level. On the other hand, social psychological research has focussed on individual psychological processes that underlie people's judgements and behaviours in carefully crafted laboratory experiments, without much consideration of the social contexts or networks in which these processes operate. However, it is clear that group-level outcomes of theoretical assumptions about intra-individual and inter-individual processes are rarely obvious, and also that individual processes often interact over time to create complex systems with non-intuitive, emergent properties (e.g. Resnick, 1994; Wolfram, 2002). A number of authors (Smith and Conrey, 2007) have therefore argued that in order to develop a full understanding of the nature of social influence, theories or models need to be constructed that take into account variables on both the individual and aggregate level of social systems. This paper introduces an attempt at such a model, by describing a connectionist Agent-based model (cABM) that incorporates detailed, micro-level understanding of social influence processes derived from laboratory studies and that aims to contextualize these processes in such a way that it becomes possible to model multidirectional, dynamic influences in extended social networks. At the micro-level, agent processes are simulated by recurrent auto-associative networks, an architecture that has a proven ability to simulate a variety of individual psychological and memory processes (Van Rooy, Van Overwalle, Vanhoomissen, Labiouse & French, 2003). At the macro-level, these individual networks are combined into a "community of networks" so that they can exchange their individual information with each other by transmitting information on the same concepts from one net to another. This essentially creates a network structure that reflects a social system in which (a collection of) nodes represent individual agents and the links between agents the mutual social influences that connect them (Hutchins & Hazlehurst, 1995; Van Overwalle & Heylighen, 2006). The network structure itself is dynamic and shaped by the interactions between the individual agents through simple processes of social adaptation. Through simulations, the cABM generates a number of novel predictions that broadly address three main issues: (1) the consequences of the interaction between multiple sources and targets of social influence (2) the dynamic development of social influence over time and (3) collective and individual opinion trajectories over time. Some of the predictions regarding individual level processes have been tested and confirmed in laboratory experiments. Additionally, data is currently being collected from real groups that will allow validating the predictions of cABM regarding aggregate outcomes.

Keywords: *connectionism, agent-based modeling, social psychology, social influence*

1. INTRODUCTION

Over the last decades, a number of researchers have started to use agent-based modeling (ABM) of collective behaviour in sociology, economics, anthropology and also psychology (Smith & Conrey, 2007). In general, ABM build social structures from the "bottom-up," by simulating individuals by virtual agents and stipulating rules that govern interactions among these agents. Creating computational models of social units (e.g. individuals, social groups, organizations or even nations) and their interactions, and observing the global structures that these interactions produce, has proven to provide unique insights into group phenomena. They express in clear mathematical and computational terms, how complex social structures emerge from computational interactions of individual agents at various distinct levels allowing the analysis of properties of individual agents (e.g. their attributes and interactions), and the emergent group-level behaviour. However, human social groups change not only through structural adaptations (i.e. social organization), but also by guiding and restructuring the behaviours and cognitions of the individuals that form them. To that extent, several modellers (Sallach, 2003; Smith & Conrey, 2007) have argued that ABM needs to incorporate relatively sophisticated models of individual agents, to allow them to adapt and change their behaviour over time. In this paper, I will introduce an ABM that aims at accomplishing that by using a recurrent connectionist model as the agent unit. It will be illustrated how such a recurrent network can simulate depersonalisation, a process that has been identified as playing a key role in social influence processes. The macro-consequences of depersonalisation are then explored by linking recurrent networks within an adaptive network structure that encodes mutual social influences within a group of agents.

2. DEPERSONALISATION

Social identity theory (SIT) and self-categorization theory (SCT) play a central role in social psychological research, and have contributed in a significant way to our understanding of the relationship between the individual and society. A central insight from this work is that individuals cognitively represent themselves in the form of self-categorizations, grouping the self and some individuals as equivalent, in contrast to other individuals. When self-categorization occurs at a group level, the *social self* or social identity is said to be salient, and the self is assimilated to other *ingroup* members – groups with which an individual identifies - and at the same time differentiated from *outgroup* members (Turner, Hogg, Oakes, Reicher & Wetherell, 1987). This cognitive redefinition of the self is called depersonalisation, or self-stereotyping in terms of an ingroup stereotype. A consequence is that individuals perceive and act in terms of their social self, rather than in terms of their personal self (Tajfel & Turner, 1986). Importantly, the term social self does not necessarily refer to demographic, sociological or role groups (e.g., women, those with low socio-economic status, or teachers). The term refers to psychological groups where an individual defines him-or herself as being a member because the group is self-relevant and self-defining. When people depersonalise, the norms, values and beliefs that define the ingroup(s) are internalised and influence the attitudes and behaviour of group members. As such, depersonalisation is seen as the main precursor to group phenomena, most notably social influence (Turner & Oakes, 1989). It is through depersonalisation that social influence becomes possible, and group processes can impact on the psychology of individual members. It results in a motivation to act in ways that advance the group's collective interests and goals and to ensure that one's own ingroup is positively distinct from other (out)groups. Because other ingroup members are viewed as similar to oneself, they become a valid source of information and a testing ground for one's own views on relevant dimensions. One's beliefs, theories and knowledge about the world and oneself are developed and validated or changed through interactions with those that are categorised as being similar to oneself (Turner, 1987). Given its central importance, it can be argued that a more thorough understanding of depersonalisation will promote our understanding of social influence processes. It is suggested here that the implementation of depersonalisation into a connectionist ABM can provide an important impetus in that process by providing a tool to model the macro-consequences of this individual psychological process in terms of socially distributed knowledge and group structures.

3. A CONNECTIONIST AGENT-BASED MODEL (cABM)

Connectionism is an approach in the fields of artificial intelligence, psychology, neuroscience and philosophy of mind, that models mental or behavioural phenomena as the emergent processes of interconnected networks of simple units. Connectionist architectures and processing mechanisms are based on analogies with properties of the human brain, in which learning is conceptualised as a process of on-line adaptation of existing knowledge to novel information provided by the environment. The focus in this paper will be on the recurrent auto-associator (McClelland and Rumelhart, 1985, 1988), a model that has been applied successfully to group biases, causal attribution & person and group impression in social psychology (Smith & DeCoster, 1998; Van Rooy et al, 2003).

3.1. Recurrent auto-associator

A recurrent network has three distinctive features (Figure 1, panel c). First, all units within an individual agent network are interconnected, such that all units send out and receive activation. Second, information is represented by *external activation*, which is automatically spread among all interconnected units within an agent in proportion to the weights of their connections. The activation coming from the other units within an agent is called the *internal activation*. Typically, activations and weights have lower and upper bounds of approximately –1 and +1. And thirdly, short-term activations are stored in long-term *weight changes* of the connections. Basically, these weight changes are driven by the difference between the internal activation received from other units in the network and the external activation received from outside sources. This difference, also called the "error", is reduced in proportion to the learning rate that determines how fast the network changes its weights and learns. This error reducing mechanism is known as the *delta algorithm* (McClelland & Rumelhart, 1985). The parameters of the individual agent nets used in the simulations below are the same as in earlier simulations by Van Rooy and colleagues (Van Rooy et al., 2003; E = D = number of internal cycles = 1, and a linear summation of internal and external activation).

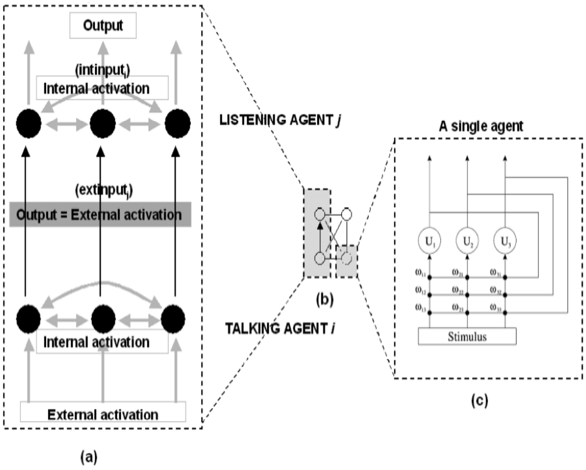

Figure 1. The left panel (a) shows the transmision of information from a talking to a listening agent. The middle panel (b) shows a group of 4 agents. The right panel (c) shows a standard recurrent network representing a single agent.

3.2. Socially distributed network & communication

A number of authors have illustrated how auto-associative networks can be naturally extended to allow communication between them (see Hutchins and Hazlehurst, 1995). It basically involves creating an agent-based model such that individual recurrent networks or *agents* are linked in an adaptive network structure. Any agent can (in principle) interact with any other agent, but the impact of the interaction will adapt to experience. Different adaptation rules have been used in previous simulations, to explore the impact of trust on communication (Van Overwalle & Heylighen, 2006) and persuasiveness of information on the development of knowledge structures (Hutchins and Hazlehurst, 1995). In the current simulation, communication involves the transmission of information from one agent's network to another, along connections whose adaptive weights reflect the mutual social influence between agents (see Figure 1, panel a). During a simulated interaction, *listening* agents compare their information (as represented by internal activation of their own network) with the information they receive from *talking* agents (represented by the external activation received from talking agents). The stronger the connection between agents, the more influence they have on each other. As such, a group of agents functions as an adaptive, socially distributed network in which information and knowledge are distributed among and propagated in function of the social influence between different individual networks. The listening agent sums all information received from other talking agents in proportion to the inter-agent weights, and then processes this information internally (according to the standard recurrent approach). Or, in mathematical terms:

$$\text{ext\_a}_j = \sum_j (w_{ij} * a_i)$$

where ext_a_j represents the external activation received by the listening agent *j*; w_{ij} is the inter-agent weight from the talking agent *i* to the listening agent *j*; and a_i denotes the final activation (which combines the external and internal activation received) expressed by the talking agent *i*.

3.3. Social adaptation

An important aspect of the cABM is that the structure in which the agents are situated is adapted through the interaction of the agents themselves. Whenever agents interact, the listening agent compares its own internal beliefs concerning an issue with the attitude expressed by a talking agent on that same issue. Inter-agent weights are then updated driven by the error between the external information, representing the attitude expressed by the talking agent, and the internal activation, representing the listening agents' attitude:

$$\delta_i = extinput_j - intinput_i$$

where $extinput_j$ is the final activation send out by the talking agent and $intinput_i$ is the internal activation of the listening agent. When agents share the same attitude, the weight of the links between them is adjusted upwards. If they disagree on an issue, the weights are adjusted downwards. This is expressed mathematically as:

$$\text{If } |ext\_a_j - int\_a_j| < \text{threshold}$$

$$\text{then } \Delta\, w_{ij} = \eta * (1 - w_{ij}) * |a_i|$$

$$\text{else } \Delta\, w_{ij} = \eta * (0 - w_{ij}) * |a_i|$$

where ext_a_j represents the external activation received (from the talking agent *i*) by the listening agent *j* and int_a_j the internal activation generated independently by the listening agent *j*; η is the rate by which the weights are adjusted. When agents largely share the same attitude (i.e. the difference is below some threshold), the links between them are strengthened. Otherwise, the links between them are weakened. This constitutes an adaptive social process, in which agents learn from interacting with each other: agents that consistently confirm each other's attitudes will be connected by stronger links than agents that consistently disagree. The social experience acquired in this way is represented in a distributed manner, in patterns of weighted links across the whole network.

4. DEPERSONALISATION AND SOCIAL INFLUENCE

I will first describe the empirical patterns associated with depersonalisation. Subsequently, a network structure is introduced that provides a cognitive mechanism that can account for these patterns. Agent networks are then embedded in an adaptive social network, to explore the interaction between depersonalisation and social interaction.

4.1. Empirical patterns

Depersonalisation has been measured in a variety of ways, including open-ended measures that request the spontaneous listing of a person's self-attributes, to asking people to judge how typical they are of a group. The consistent finding is that it reconfigures our representation of ourselves to conform to the prototype of an ingroup, such that the self is viewed through the lens of the relevant ingroup (Turner, 1987) and is predominantly described in terms of traits or characteristics of that ingroup, rather than distinctive, individuating traits. For instance, individuals will describe themselves as being more typical of an ingroup and less typical of an outgroup when depersonalised, as compared to when they perceive themselves as individuals. In addition, a number of studies seem to suggest that depersonalised individuals are more open to influence during social interaction. For instance, groups of individuals show more consensus in shared knowledge structures, such as stereotypes, when their interactions are framed in terms of a shared, social identity (Haslam, Turner, Oakes, Reynolds, 1998). In the following simulation, we will focus on these 2 patterns: (1) depersonalisation leads to more perceived similarity between self and ingroup, and more perceived differences with an outgroup; (2) social interaction produces more consensus in stereotypes when it is predicated on a shared social self.

4.2. Theoretical assumptions and network structure

At present, there is no detailed model of the mental representations or processes that might underlie the findings in this area. Instead, depersonalisation has been typically explained in terms of a metaphorical merging of *self* and *other* (i.e. an ingroup or outgroup member) representations. The connectionist approach makes the explicit assumption that all mental representations are encoded and interconnected within the same network, and that contextual cues determine which self-categorization takes place. This is reflected in the agent network structure in Figure 2: representations of self and group stereotypes are distributed patterns of activations across a number of *trait* nodes. Because of its simplicity and ease of interpretability, a localist encoding is used, where each node represents a specific trait.

However, there is no single node representing group membership or self per se, rather we assume that parts of the distributed pattern represent configurations of traits that are apparent through self-perception as cues to the personal self, whereas other parts represent traits that are perceived to be correlated with group membership. If the context primes information highly associated with the *personal* (i.e. unique traits) but not the social self (i.e. ingroup traits), representations close to self will be more strongly activated then those close to the ingroup. Conversely, when, through the same process, characteristics of the ingroup dominate, self-categorization occurs more in terms of the ingroup stereotype. This is essentially a socially situated approach to cognition, where the social context – which can represent a real-life social situation, an experimental lab situation, or specific questions or manipulations by the experimenter – determines whether an individual self-categorizes mainly in terms of an ingroup stereotype ("I am a typical student") or in terms of a personal self ("I am not a typical student").

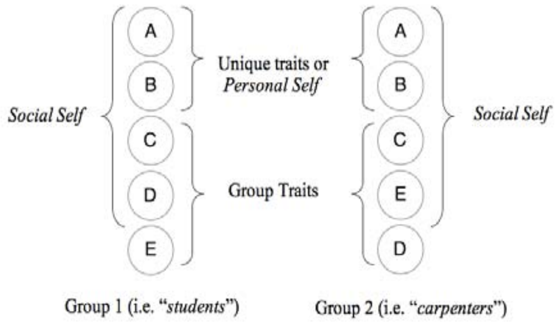

Figure 2. Agent networks representing a personal self, ingroup, outgroup and social self stereotype

To achieve the network structures in Figure 2, a population of networks was trained with a series of patterns with information about the relationship between 5 traits ("ABCDE"). A set of patterns was constructed that associated attribute A & B, simulating a group of agents that define themselves mainly in terms of some configuration of these 2 attributes. This represents the *personal* self of these networks. Another set of patterns was constructed that associated this personal self with either attributes CD *or* CE. These associations define group membership, and networks were trained in such a way that they were either associated with group 1 (attributes CD) or 2 (attributes CE). As such, the associations between ABCD, or ABCE, represent the *social* self of a network. Each network thus learns 4 stereotypes: a personal & social self, and ingroup and outgroup stereotype. The social self includes both the personal self and ingroup stereotypes, which either consists out of traits ABC*D* or ABC*E*, depending on the group association. Psychologically, this corresponds to a situation in which members from 2 social groups develop stereotypical impressions of themselves and their groups. Through direct experiences (observations of self and others) and indirect experiences (communication or observation of others' experiences), they develop expectancies about which traits characterise themselves and their ingroup, and how they set themselves and their groups apart from others. A group of students define themselves in terms of two unique attributes (AB), but also in terms of *diligence* (C) and *intelligence* (D) that are shared by many students in varying degrees (i.e. not all students are equally intelligent, but as a group they are more intelligent than carpenters). Similarly, a group of carpenters might define themselves in terms of a particular configuration of unique attributes (AB), but also in terms of *diligence* (C) and *independence* (E), but not so much *intelligence* (D).

4.3. Testing the networks

We simulate a self-categorization measure by testing our agent networks with 2 different cues or probes: (1) A *personal* self probe, in which both unique traits (AB) are maximally activated (i.e. a series of + 1.0); (2) a *social* self probe, in which one unique (A) and one group trait (C) are activated. Activation values are then allowed to flow through the network, and the extent to which the network activates the ingroup and outgroup nodes (either attribute C or E, depending on the group association) indicates the strength of association between an agent and these groups. Psychologically, this would correspond to asking an individual how typical she is of a particular group. Figure 3 shows average simulated stereotypicality judgments for ingroup and outgroup in function of probe type. As would be expected, agent networks categorise themselves as

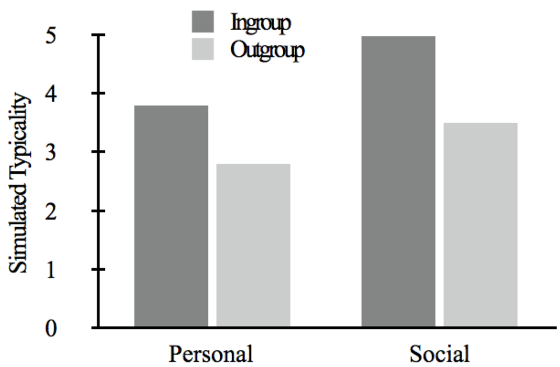

Figure 3. Average simulated stereotypicality judgement in terms of ingroup and outgroup, and in function of probe type. Higher bars indicate more stereotypical judgments.

more stereotypical for ingroup than outgroup. Importantly, the figure shows that this is more so when networks are tested with a social probe. In other words, if the context primes information highly associated with the social self, self-categorization occurs more in terms of the ingroup stereotypy, leading to more depersonalisation. Figure 3 also shows that the difference between ingroup and outgroup stereotypicality is larger in the social as compared to the personal condition, reproducing the finding that differences between self and outgroup members are emphasised when an individual depersonalises.

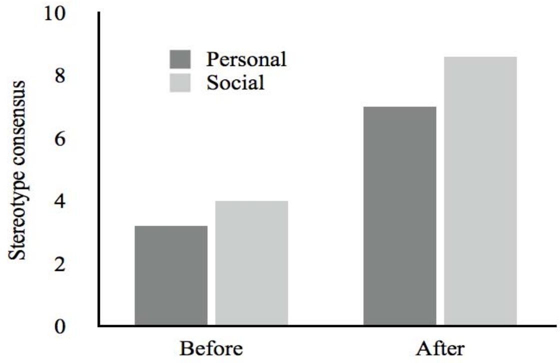

Figure 4. Average simulated stereotype consensus in function social interaction. Higher bars indicate more concensus amongst agents.

As mentioned, a number of studies have shown that social interaction produces more consensus in stereotypes when it is predicated on a shared social self or identity. Our agent-based implementation allows us to explore this process by simulating interaction between 2 groups of agents that involves a talking and listening phase during which all agents communicate with each other. Figure 4 shows stereotype consensus within each group of agents both *Before* and *After* simulated social interaction, and captures the finding that social interaction enhances consensus in ingroup stereotypes, and also that this effect is larger when individuals depersonalise in the social probe condition (Haslam et al, 1999).

4.4. Agent opinion trajectories

Because the model includes representations of individual cognitions (agent recurrent networks), it becomes possible to analyse how information that is communicated through the social system is adapted and integrated. Each time an agent acquires information, it assimilates and adds its own personal experience (as captured by the long-term weights within an agent network) before sending its' interpretation of the received information out again into the group. One can think of this as a game of Chinese whispers: Every member of the communication chain adds his or her own interpretation to the information, leading to changes as the information proceeds down the communication chain. It is through this process that agents and the information they hold undergo a process of self-organization, whereby out of local interactions global, more consensual structures emerge.

Figure 5 shows how agent opinion trajectories develop over the course of simulated social interaction for 2 groups of agents. Position in opinion space was determined by testing each network with a social self probe and measuring the extent to which it filled in the ingroup attribute (either E or D, depending on group membership). Networks that produce similar outputs are closer together, which conceptually represents similar categorisation in terms of an ingroup stereotype. After each measurement networks interacted, after which they were probed again. The figure shows that, as social interaction unfolds, the distance between the 2 groups becomes larger, illustrating how individual agents self-organise in clusters of attitudinal similarity. The links within these clusters grow stronger, while links between them grow weaker. Even though the set-up of this simulation is relatively simple, the behaviour of the agents shows remarkable similarities to well-known social psychological processes: Agents organise themselves in clusters of agents that either agree (the ingroup) or disagree (the outgroup) on certain issues, and stronger connection weights between similar agents

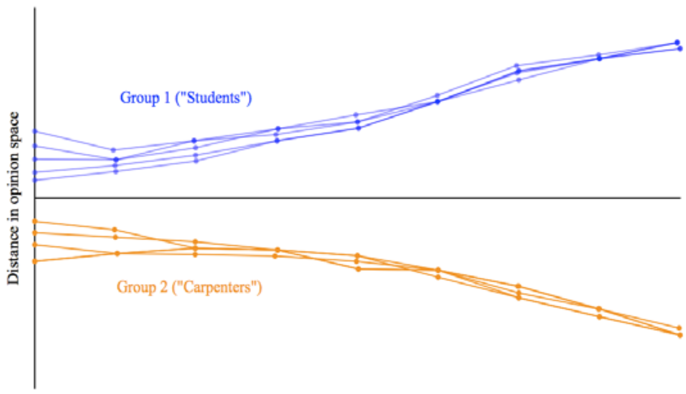

Figure 5. The trajectories of 2 groups of 5 agents (i.e. “students” and “carpenters”) in function of social interaction. Greater disctance in opinion space represents greater difference between ingroup stereotypes.

reflect increased social influence within such clusters. The simulations show how the strengthening links between agents within a single sub-cluster act as positive feedback loops that result in agents reinforcing each other's attitudes. This essentially leads to a group polarisation effect (as apparent in the increasing distance between opinion clusters), as the agents end up with more extreme opinions after the interaction, and also more consensual ingroup stereotypes. This simulated process thus shows strong similarities with the process through which real social groups create, validate and maintain socially shared knowledge, and mimics how group membership attenuates social influence: Agents are more likely to conform to other agents within the same cluster (the ingroup), because of the high mutual social influences within that cluster (Sherif, 1935; Turner et al., 1987).

5. DISCUSSION AND CONCLUSIONS

The objects of psychological inquiry are complex systems that afford analysis at different levels of description. Our understanding of a given phenomenon gains explanatory power particularly when we can provide a causal account of it in terms of the entities and organising principles at a lower level of description than the phenomenon itself (Marr, 1982). Connectionist principles are cast at a lower level of description than the level of description that is appropriate to describe their behaviour, and bear no transparent relationship with the phenomena that they are able to account for (*i.e. depersonalisation, social influence*). There were current theorising in psychology is very much couched in verbal, theoretical descriptions, the connectionist perspective provides an account for complex social categorisation processes based on very simple, but powerful algorithms that mimic real memory processes. By developing a cognitive agent that implements basic self-categorization processes in terms of connectionist principles, and embedding such an agent within an adaptive network structure, we can start exploring macro level-consequences of the repeated application of these processes, in parallel by many agents, within an artificial social system. Such an integrated framework will allow investigating the interaction between memory (i.e. pattern learning and retrieval), individual (i.e. depersonalisation) and group (social influence, communication) processes in fundamentally novel ways.

REFERENCES

Haslam, S. A., Oakes, P. J., Reynolds, K. J., & Turner, J. C. (1999). Social identity salience and the emergence of stereotype consensus. *Personality and Social Psychology Bulletin, 25*(7), 809-818.

Hazlehurst, B., & Hutchins, E. (1998). The emergence of propositions from the co-ordination of talk and action in a shared world. *Language and Cognitive Processes*, *13*, 373-424.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology, 114*, 159–188. Library Holdings

McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs and exercises*. Cambridge, MA: Bradford.

Resnick, M. (1994). *Turtles, termites, and traffic jams: Explorations in massively parallel microworlds.* Cambridge, MA: MIT Press.

Sallach, D. (2003). Social theory and agent architectures: Prospective issues in rapid-discovery social science. *Social Science Computer Review, 21*, 179-195.

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology, 27*, 1–60.

Smith, E. R., & Conrey, F. R. (2007). Agent-based modeling: A new approach for theory-building in social psychology. *Personality and Social Psychology Review, 11*, 87-104

Smith, E. R., & DeCoster, J. (1998). Knowledge acquisition, accessibility, and use in person perception and stereotyping: Simulation with a recurrent connectionist network. *Journal of Personality and Social Psychology, 74*, 21–35.

Tajfel, H. and Turner, J. C. (1986). The social identity theory of inter-group behavior. In S. Worchel and L. W. Austin (eds.), *Psychology of Intergroup Relations*. Chicago: Nelson-Hall.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory.* Cambridge, MA: Basil Blackwell.

Turner, J. C., & Oakes, P. J. (1989). Self-categorization theory and social influence. In P. B. Paulus (Ed.), *The psychology of group influence*, Hillsdale, NJ: Lawrence Erlbaum.

Van Rooy, D., Van Overwalle, F., Vanhoomissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review, 110*, 536-563.

Wolfram, S. (2002). *A new kind of science*. Champaign, IL: Wolfram Media.