This paper introduces 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 [1]. 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 [2]. 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. In an extensive research program, data is currently being collected from real groups that will allow validating the predictions of cABM regarding aggregate outcomes.

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

1. Introduction

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 behaviors 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 [3, 4]. A number of authors [5, 6] 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.

A potential method to achieve this is the use of agent-based modeling (ABM). In general, ABM build social structures from the "bottom-up", by simulating individuals with 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 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 behavior. However, human social groups change not only through structural adaptations (i.e. social organization), but also by guiding and restructuring the behaviors and cognitions of the individuals that form them. To that extent, several modelers [7, 8] have argued that ABM needs to incorporate relatively sophisticated models of individual agents, to allow them to adapt and change their behavior over time.

In this paper, an ABM is introduced that aims at accomplishing this by implementing self-categorization processes into a recurrent, connectionist agent model. I hope to illustrate how the use of a valid, social psychological theory of agent heuristics can contribute to a better understanding of social complex phenomena at the macro level. First, the model is used to simulate a number of key empirical patterns from the self-categorization literature. Subsequently, the predictions of the model are tested in an empirical study, in which the emergence of shared social categories from inter and intra personal processes was investigated. And finally, the cABM is compared to similar models, and it is discussed how it can be used to further explore the interaction between the individual and aggregate levels of social systems.

2. Personal and social self

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 out-group members [9]. This cognitive redefinition of the self is called depersonalization, 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 [10]. 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 depersonalize, the norms, values and beliefs that define the ingroup(s) are internalized and influence the attitudes and behavior of group members. As such, depersonalization is seen as the main precursor to group phenomena, most notably social influence [11]. It is through depersonalization 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 categorized as being similar to oneself.

Given its central importance, it can be argued that a more thorough understanding of self-categorization in terms of a personal or social self will promote our understanding of social influence processes. It is suggested here that the implementation of self-
categorization into a connectionist ABM (or cABM) can provide an important impetus in that process, in two major ways. First, it can provide a more detailed, cognitive agent-model of individual self-categorization processes then is currently available. Secondly, most of our knowledge on self-categorization comes from laboratory studies, in which social influence is manipulated and analyzed at the level of individual cognitive processes. However, as pointed out by a number of authors (see [12]), simple, individual processes often combine to create complex systems with nonintuitive emergent properties when they are iterated across time and space. A cABM of self-categorization allows exploring such emergent properties in more detail than current, mostly verbal theories and models, for instance by generating detailed predictions about the dynamic development of reciprocal social influences within networks agents.

3. The 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 behavioral 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 conceptualized 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 [13, 14], a model that has been applied successfully to group biases, causal attribution & person and group impression in social psychology [1, 15].

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
Connectionist ABM

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.

In a recurrent network, processing information takes place in two phases. During the first phase, each unit in the network receives activation from external sources. Because the units are interconnected, this activation is automatically spread throughout the network in proportion to the weights of the connections to the other units. The activation coming from the other units is called the internal activation (for each unit, it is calculated by summing all activations arriving at that unit). Together with the external activation, this internal activation determines the final pattern of activation of the units (termed the net activation), which reflects the short-term memory of the network. In the linear version of activation spreading in the recurrent network that is used here, the final activation is the linear sum of the external and internal input after a single updating cycle through the network. In nonlinear versions used by other researchers (see for instance [13]), the final activation is determined by a nonlinear combination of external and internal inputs updated during a number of internal cycles (for mathematical details, see [14]). Previous simulations by Van Rooy and colleagues revealed that the linear version with a single internal cycle reproduced observed data at least as well, suggesting that the present linear activation update algorithm with a single internal cycle is sufficient for simulating many phenomena in group judgments (see for instance Table 6 in [1]).

The net activation of a unit is determined by the sum of the external and internal activations, after one updating cycle through the network. More specifically, every unit $i$ in the network receives external activation, termed $ext_i$, which is the sum of the activation from the other

$$a_i = E \times ext_i \quad (1)$$

This activation subsequently spreads through the auto-associative network, meaning every unit $i$ receives internal activation $int_i$, which is the sum of the activation from the other units $j$ (denoted by $a_j$) in proportion to the weight of their connection to unit $i$, or for all $j \neq i$.

$$int_i = \sum (w_{ji} \times a_j) \quad (2)$$

External and internal activation are then summed to the net activation, or

$$net_i = E \times (ext_i + int_i) \quad (3)$$

The updating of activation at each cycle is governed by the following equation [14]:

$$\Delta a_i = net_i - D \times a_i \quad (4a)$$

where $D$ reflects a memory decay term. As in previous simulations [1], parameter values were set to $D = E = 1$. Hence, the final activation of unit equals the sum of the external and internal activation, or

$$a_i = net_i = ext_i + int_i \quad (4b)$$
After activation has been determined, the recurrent model enters the second learning phase in which the 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 [14]. In mathematical terms, the delta algorithm strives to match the internal predictions of the network int<sub>i</sub> as closely as possible to the actual state of the external environment ext<sub>i</sub> and stores this information in the connection weights. This error-reducing process is formally expressed as [14]

$$\Delta w_{ji} = \varepsilon \times (ext_i - int_i) \times a_j$$

where $\Delta w_{ji}$ is the change in the weight of the connection from unit j to i, and $\varepsilon$ is a learning rate that determines how fast the network learns. An implication of this learning algorithm is that when an object and its feature co-occur frequently, then their connection weight gradually increases to eventually reach an asymptotic value of +1.

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 [2, 16]). 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 [16] and persuasiveness of information on the development of knowledge structures [2]. 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:

$$ext_{-i}a_i = \sum_j w_{ki} \ast a_i$$

where $ext_{-i}a_i$ represents the external activation received by the listening agent $l$; $w_{ki}$ is the inter-agent weight from the talking agent $k$ to the listening agent $l$; 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:

\[ \delta_i = \text{extinput}_j - \text{intinput}_i, \]

where \( \text{extinput}_j \) is the final activation send out by the talking agent and \( \text{intinput}_i \) 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:

\[
\begin{align*}
\text{If } |\text{ext}_a - \text{int}_a| &< \text{Tolerance} \\
\text{then } \Delta w_{kl} &= \eta \times (1 - w_{kl}) \times |a_i| \\
\text{else } \Delta w_{kl} &= \eta \times (0 - w_{kl}) \times |a_i|
\end{align*}
\]

are adjusted downwards. This is expressed mathematically as:

where \( \text{ext}_a \) represents the external activation received (from the talking agent \( k \)) by the listening agent \( l \) and \( \text{int}_a \) the internal activation generated independently by the listening agent \( l \); \( \eta \) is the rate by which the weights are adjusted. When agents largely share the same attitude (i.e. the difference is below the Tolerance 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.

3.4. Self-organization

The cABM models social groups as non-linear, dynamical systems, in that we expect group and shared knowledge structures to emerge from social interaction through self-organization of the constituent elements. If the individual agents within the network react in a consistent and coherent way to the information they receive from each other (in the form of activation spreading within the system), the cumulative effect of their local adjustments to their immediate environment will result in coordinated patterns at the group level. Connectionist systems lend themselves perfectly for modeling this type of self-organization process. Within the connectionist approach, both groups and individuals are seen as complex, dynamical systems that can be characterized by the connections between the constituent elements. Applied to a group, the connections represent communication channels that allow the transmission of information and, through their adaptive nature, lead to coordinated action at the group level. These networks have no central executive, but instead adapt through simple, local algorithms that adjust the connections between the constituent elements of the system. Also, connectionist systems are known to evolve into local attractor states. These are patterns of activation that are more likely
than others, in terms of the current information presented to a system and all prior information embedded within the connections of that system. Through the repeated application of local learning algorithms, a connectionist network can be seen as self-organizing by moving towards an attractor state. Within the current framework, these attractor states are considered to be consensual group structures that incorporate all possible constraints within the system.

4. Stereotyping and social influence

I will first illustrate how the cABM can fit critical empirical patterns associated with stereotyping, self-categorization and social influence. 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 impact of social interaction.

4.1. Empirical patterns

As described in section 2, self-categorization theory essentially states that the perception an individual has of him or herself is context-dependent. In many contexts, self-perception will focus on individual characteristics (i.e. “academic”, “spouse”, “musician”) that distinguish us from other individuals, providing us with a personal identity. Self-categorization at a group level occurs when a social identity is salient, which is defined as “[... ] that part of the individuals' self-concept which derives from their knowledge of their membership of a social group (or groups) together with the value and emotional significance attached to that membership.” [10]. When that occurs, an individual will self-stereotype in terms of an ingroup, which is a group with which an individual identifies, and which is differentiated from out-groups. For example, a group of Europeans is more likely to categorize themselves as 'European' (rather than as 'French' and 'Germans', say) in a situation where Americans and Asians are also present, rather than just other Europeans. This cognitive redefinition of the self is called depersonalization, or self-stereotyping in terms of an ingroup stereotype. This phenomenon 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 when the social self is salient, individuals reconfigure their self representation to conform to the prototype of an ingroup, such that the self is viewed through the lens of the relevant ingroup and is predominantly described in terms of traits or characteristics of that ingroup, rather than distinctive, individuating traits. Under those circumstances, individuals describe themselves as being more typical of an ingroup and less typical of an outgroup, as compared to when the personal self is salient. In addition, a number of studies seem to suggest that when the social self is salient, individuals are more open to social influence. 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 [17]. In the following simulation, we will illustrate how the cABM provides a novel, alternative explanation of these 2 critical patterns: (1) self-categorization to a social self leads to more perceived similarity between self and ingroup, and more perceived differences with an outgroup; (2) social interaction predicated on that social self produces more consensus in stereotypes.
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, self-categorization 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.

![Diagram of agent networks representing a personal self, ingroup, outgroup and social self stereotype](image)

There is no single node representing group membership or (personal or social) 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 (personal or social) self will be more strongly activated than 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”).

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). 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 characterize themselves and their ingroup, and how they set themselves and their groups apart from others. For instance, 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 might believe that 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 stereotypical judgments for ingroup and outgroup in function of probe type. As would be expected, agent networks categorize themselves as 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 self-categorization. Figure 3 also shows that the difference between ingroup and outgroup stereotypicality is larger in the social compared to the personal condition, reproducing the finding that differences between self and outgroup members are emphasized when an individual “depersonalizes” in terms of a social self.

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. An interaction involves a talking (or sending) and listening (or receiving) agent, and is completely determined by the equations introduced earlier. More specifically, external activation is provided to a talking agent (1), which generates internal activation within that agent (2, 4b). Communication then involves a listening agent receiving activation form a talking agent (6), which then generates internal activation within the listening agent (2, 4b). After the interaction, social adaptation takes place (8), and then individual agents learn as well (5).
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 de-personalize in the social probe condition [17].

4.4. Agent stereotype trajectories

Because the model includes representations of individual cognitions (agent recurrent networks), it becomes possible to analyze 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 ingroup stereotypes develop over the course of simulated social interaction for 2 groups of agents. Each agent position in the graph was determined by testing a 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 categorization 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-organize in clusters of stereotypical 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 behavior of the agents shows remarkable similarities to well-known social psychological processes: Agents organize themselves in clusters of agents that either agree (the ingroup) or disagree (the outgroup) on certain issues, and stronger connection weights between similar agents 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 polarization 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 [5, 6].
4.5. Dynamic network within and between agents.

Before moving onto the empirical test of the model, I will use an example based on the design of the study below to illustrate the dynamic aspect of the connection weights both within and between agents. As mentioned, the weights of the connections between agents determine the influence agents have on each other. These weights are subjected to learning and change in almost exactly the same way as the internal weights of agent networks. Whereas internal weights encode an individual agents’ learning history, the external connection weights encode the history of interactions between agents. Let’s take a simple interaction between 2 agents comparing information they have on a fictional “Group A” (see Figure 6).

Fig. 6. The adjustment of inter-agent weights in communication between agents. Here, a talking agent expresses its opinion on the “Creativity” of the members of “Group A”. Because the expressed opinion of the talker and the private opinion of the listener are identical, it will lead to an increase of the relevant connection weights.

Suppose both agents agree that members of Group A are relatively creative, which is encoded in the network of both agents by similar connection weights (.6) between the unit representing “Group A” and the unit representing the attribute “Creative”. An interaction between a talking (or sending) and listening (or receiving) agent proceeds according to the equations introduced earlier: First, the “Group A” unit in the talker network is activated, by providing an activation of +1. This activation spreads internally to the “Creative” unit in proportion to the connection weight (.6) and results in an activation of .6 of that unit. When the Talking agent expresses its impression of Group A to the Listening agent, the activation of the “Group A” unit is transmitted to the Listener along the relevant external connection, which initially has a default weight of .5. As a result, the “Group A” unit in the listening network is activated (.5), and this activation spreads internally to the “Creative” unit in proportion to the connection weight (.6) and results in an activation of .3 of that unit. This internal activation is then compared to the corresponding external activation received from the Talking agent. More precisely, the Creative unit within the talking network generates internal activation of .6, which is transmitted as external activation to the Listener network, in proportion to the weight of the external connection (.5). The resulting external activating arriving at the Listener is .3. In this case, it means the communication by the talking agent and the opinion of the listening agent are similar (.3). As a result, the connection weight between the Creative unit in the talking
and listening networks will increase in strength according to formula 8. However, as soon as the difference between external and internal activation exceeds the Tolerance threshold, that weight will begin the decrease. As such, the relative agreement that agents develop during interaction on particular topics, will be represented in the weights that externally connect these topics between agents.

5. Tolerance and Need for Closure

Despite its central importance, only a few studies [17] have investigated the actual development of consensual categories within real, interacting groups of participants. This is partly due to the reluctance of psychological research to use the group as unit of analysis instead of the individual, but also to the complexity of the subject. Interactions between even small groups of individuals result in complex patterns of reciprocal influences, in which actions of one individual can have non-linear implications for the whole social system. Especially given the fluid nature of social categories, it would be impossible to derive precise predictions using existing verbal theories about the development of socially shared categories in the course of social interaction. The use of a computational model alleviates these problems, and allows making predictions of how processes working on different levels of a social system interact to produce emergent properties.

5.1. Small group study

Psychological research has identified a number of socio-cognitive variables that reflect individual differences in the willingness to deal with ambiguity in the social environment. For instance, Need for Closure (NfC) reflects a desire for a quick and definitive answer to any question or decision rather than sustained uncertainty, confusion, or ambiguity [18]. We set out to test to what extend the Tolerance parameter in our model (see section 3.3) can simulate variation in social behavior caused by NfC, both in small group experiments and in larger collectives. In cABM, the Tolerance parameter indicates the amount of difference in attitudes agents will allow before the links between them are weakened. Our hypothesis was that high tolerance (agents accept large differences in attitudes) would correspond to low NfC, and vice versa. This hypothesis was tested in a small group study.

5.1.1. Subjects

One hundred and fifty psychology undergraduate students (40 men, 110 women; mean age = 22.63) participated in the study as part of a class exercise. Prior to the actual experiment, all participants filled out a dispositional Need for Closure scale [18] in an apparently unrelated session.

5.1.2. Procedure

Participants arrived in the lab and were informed that they would be receiving information about individuals who belonged to one of two groups (Group A & B). They were asked to form an impression of these groups, and told that they would afterwards share their impressions with other participants. Based on their NfC score, participants were assigned in groups of three to either a High or Low NfC condition. Individual participants were then presented with written information describing members of the two groups in the form of statements (For instance, “I am finishing up my medical degree at Johns Hopkins in cardio-thoracic surgery. Before this, I received my doctorate from Cambridge in molecular biology. After working in the field, I decided to attend medical
school to become a surgeon.”). Crucially, some of the information about the groups was available to all group members (*shared*) whereas the remainder was distributed among group members (*unique*). More particularly, for both group A & B there were 4 critical attributes distributed across groups of three participants, so that each participant had one unique piece of information, and one piece of information was distributed across the three participants (see Table 1). For instance, participant 1 would learn that members of group A are sociable and intelligent, while participant 2 received information indicating that members of group A were sociable and creative. Because we wanted to study how the impressions of the group evolved during social interaction, this initial information did not associate either group strongly with any attribute. Participants read this information individually, and then indicated their impressions of the groups on a range of dependent measures.

### Table 1: Distribution of information across participants during individual phase

<table>
<thead>
<tr>
<th>Type of information:</th>
<th>Shared</th>
<th>Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>D</td>
</tr>
</tbody>
</table>

*Note.* Information pertaining to 4 different traits (A, B, C, D) is distributed amongst groups of 3 participants, such that one trait (A) is shared, while the others (B-D) are “unique” to individual participants.

Subsequently, participants were put together in groups of 3. An experimenter lead the group session. Each trial involved the experimenter reading out a statement describing the behavior of an individual (e.g. “We feel good when we are in the company of others” as an example of a sociable behavior) and then asking each participant for their judgment. Participants were instructed to indicate on a rating scale the degree to which they considered each statement to be representative of group A and B, and to read aloud the number they assigned (i.e. “Seven”). The order in which participants answered was randomized across trials. Twelve statements were presented for each critical attribute, totaling 48 statements per group. At the end of the group interaction phase, participants were separated and completed the same dependent measures they completed at the start of the procedure. On finishing the study, participants were debriefed and thanked for their participation.

### 5.1.3. Data

The basic idea behind this controlled methodology is to encourage individuals to learn from the classification behavior of others. At the end of the procedure, the categorization behavior of each individual group member can be conveniently represented by vectors. For instance, the perceived relationship between the attribute “friendly” and group A was assessed on 12 different occasions (i.e. participants rated on 12 occasions how typical someone behaving in a friendly manner was of group A). For each participant, this results in a 12 dimensional vector, where the separate components correspond to 12 different ratings given by that participant. By comparing these individual vectors within groups, we are able to quantify the level of variance or consensus within each group.
5.1.4. **Hypotheses**

The 2 top panels of Figure 7 show how in both NFC conditions, the cABM predicts that the variance in categorization within a group will become smaller, while stereotype strength will increase over the course of the experiment. The cABM also predicts that both consensualization and strengthening of the stereotype will be more outspoken in the Low NFC condition. The bottom panels show that results largely confirmed cABM predictions: The left panel shows that the reduction in variance corresponded to the predicted trend (Model fit: $r = .8$, $p < .01$). Similarly, the cABM prediction regarding the strength of the stereotype was also confirmed, even though the model underestimated the intercept (Model fit: $r = .72$, $p < .01$): As the experiment progressed, participants gradually reinforced each others categorizations, leading to essentially a polarization effect (i.e., higher ratings on the scale). As predicted, both effects were more outspoken in the Low NFC condition.
5.2. Further research

By creating a community of networks, new parameters appear that are not present when we only consider an individual network. Table 1 lists cABM parameters, and their hypothesized social psychological role. It is the nature of these parameters, and the nature of their relationship with features of real social groups, that are the focus of our empirical program. Small group studies, like the one described above, allow us to validate and inform the most psychologically plausible values of these parameters. For instance, our studies have shown that the variation in social behavior caused by individual differences in Need for Closure can be successfully simulated by allowing the Tolerance parameter to vary between approximately .25 (low tolerance, corresponding to high NfC) and .4 (high tolerance, corresponding to low NfC). This provides us with a psychologically plausible range of values, which can be used to inform simulations of larger collectives, such as groups of employees in organizations.

Although not addressed in the current paper, the cABM generates a number of predictions about how dissonance between multi-dimensional attitudes affects the development of consensus. Surprisingly, this topic has received relatively little attention in social psychological research, but there are a number of simulation models that have explored it. For instance, Huet and colleagues [19] demonstrated the impact of multi-dimensional attitudes on the development of conformity, using a model very similar to the cABM: Both models are build around a rejection mechanism determined by a tolerance parameter, and both generate aggregate predictions consistent with self-categorization theory. The experimental design introduced in this paper could be adapted, for instance by providing information to participants that is inherently contradictory (i.e. Group A is both aggressive and peaceful). This would allow testing the predictions of the cABM, and to formally compare it with other simulations.

In general, relevant psychological theories can be instantiated in cABM in a way that supports a process of probing and prediction. In this process, simulations provide guidance for empirical research as well as sufficient depth to support interactive modification of the underlying theory [1]. Using a computational model, we can create large sets of simulated social groups, each set having its own characteristics, which can then be put into a wide range of conditions. This represents a very efficient way of testing theoretical predictions in a simulated collective. Using this approach, we are currently exploring how large groups of employees within organizations develop a shared organizational identity, how that relates to a number of critical socio-cognitive variables (for instance NfC), but also how the amount of social interaction, or the types of networks within organizations, affect this process.

6. Comparison with other models.

The study of social influence has received a substantial amount of attention in several disciplines. The cABM shares similarities with some of this work, and in particular with models that postulate mixes of assimilative and contrastive social influence. Much like the cABM, a number of models use a dynamic matrix of weights to represent how reciprocal social influence between individuals evolves over time [16, 20]. These models posit that individuals accept influence only from others whose current opinions are within a certain threshold distance from their own current positions. Similarly, the class of “se-
cABM models postulate that people are motivated to seek distinctiveness when they are in company with others. Typically, every individual is assigned a particular attitude posi-

tion on a continuous scale, and then looks at randomly chosen others and selects the one whose attitude is the most different from the mean attitude within a particular comparison group. All of these models are very consistent with self-categorization theory, and thus cABM. They all predict that individuals will move to the position of a deviant, noncon-
forming position but only in company with at least a few similar others. As such they combine an assimilative process, i.e. moving toward a chosen individual, with a contrastive process, moving toward the extreme and away from the mean. Although formal sim-
ulations are needed, it is likely that the predictions of these models of how groups can end up converging to a common opinion or split into several subgroups holding differing opinions, depending on the initial attitude distribution and the threshold for influence from others, are very similar.

7. Conclusion

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 organizing principles at a lower level of description than the phenomenon itself. Connectionist principles are cast at a lower level of description than the level of description that is appropriate to describe their behavior, and bear no transparent relationship with the pheno-
mena that they are able to account for (i.e. self-categorization, social influence). There
were current theorizing in psychology is very much couched in verbal, theoretical de-
scriptions, the connectionist perspective provides an account for complex social categori-
ization 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 re-

<table>
<thead>
<tr>
<th>Agent</th>
<th>Individual Recurrent Network</th>
<th>Knowledge structures</th>
<th>Collective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local dynamics (Agent level)</td>
<td>Intragroup characteristics</td>
<td>Social identity</td>
<td>Social group, organization, ...</td>
</tr>
<tr>
<td>1. Pattern of connections within agent</td>
<td>1. Schemata for phenomena (i.e. background theories)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Pattern of activation within agent network</td>
<td>2. Current individual opinion, attitude, belief or attributes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. External input to agent network</td>
<td>3. Access to environment (context, information)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global dynamics (Group level)</td>
<td>Group characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Patterns of connections between agents</td>
<td>5. Communication structures, relations within group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Connection weights between agents</td>
<td>6. Identity-based and normative factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. External input to group of agents</td>
<td>7. Distribution of information in small group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Number of iterations</td>
<td>8. Time (i.e. stage of group development)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Social Adaptation rule</td>
<td>9. Reinforcement of consensus, diversity or a mixture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Inter-agent Weight Change Rate</td>
<td>10. Flexibility of group relations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context dynamics</td>
<td>Context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affect all of the above</td>
<td>Group interdependence (competition, discrimination,...)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
peated 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. self-categorization) and group (social influence, communication) processes in fundamentally novel ways.

References


