Talking Nets: A Multi-Agent Connectionist Approach to Communication and Trust between Individuals

Frank Van Overwalle & Francis Heylighen

Vrije Universiteit Brussel, Belgium

This research was supported by Grant OZR941 of the Vrije Universiteit Brussel to Frank Van Overwalle. We are grateful to Margeret Heath for her suggestions on an earlier draft of this manuscript. Address for correspondence: Frank Van Overwalle, Department of Psychology, Vrije Universiteit Brussel, Pleinlaan 2, B - 1050 Brussel, Belgium; or by e-mail: Frank.VanOverwalle@vub.ac.be.

Abstract

A multi-agent connectionist model is proposed that consists of a collection of individual recurrent networks that communicate with each other, and as such is a network of networks. The individual recurrent networks simulate the process of information uptake, integration and memorization within individual agents, while the communication of beliefs and opinions between agents is propagated along connections between the individual networks. A crucial aspect in belief updating based on information from other agents is the trust in the information provided. In the model, trust is determined by the consistency with the receiving agents' existing beliefs, and results in changes of the connections between individual networks, called trust weights. Thus activation spreading and weight change between individual networks is analogous to standard connectionist processes, although trust weights take a specific function. Specifically, they lead to a selective propagation and thus filtering out of less reliable information, and they implement Grice's (1975) maxims of quality and quantity in communication. The unique contribution of communicative mechanisms beyond intra-personal processing of individual networks was explored in simulations of key phenomena involving persuasive communication and polarization, lexical acquisition, spreading of stereotypes and rumors, and a lack of sharing unique information in group decisions.

Cognition is not limited to the mind of an individual agent, but involves interactions with other minds. A full understanding of human thinking thus requires insight into its social development. Sociologists and developmental psychologists have long noted that most of our knowledge of reality is the result of communication and social construction rather than of individual observation. Moreover, important real-world problems are often solved by a group of communicating individuals who pool their individual expertise while forming a collective decision that is supposedly better than what they might have achieved individually. This phenomenon may be labeled *collective intelligence* (Levy, 1997; Heylighen, 1999) or *distributed cognition* (Hutchins, 1995). To understand such group-level information processing, we must consider the distributed organization constituted by different individuals with different forms of knowledge and experience together with the social network that connects them and that determines which information is exchanged with whom. In addition to qualitative observations and conceptualizations, the phenomenon of distributed cognition has been studied in a quantitative, operational manner, although from two very different traditions: *social psychology* and *multi-agent simulations*.

Social psychologists have typically studied group level cognition using laboratory experiments, and their research has documented various fundamental shortcomings of collective intelligence. We often fall prey to biases and simplistic stereotypes about groups, and many of these distortions are emergent properties of the cognitive dynamics in interacting minds. Examples are *conformity* and *polarization* which move a group as a whole towards more extreme opinions (Ebbesen & Bowers, 1974; Mackie & Cooper, 1984; Isenberg, 1986), communication within groups which reinforce *stereotypes* (e.g., Lyons & Kashima, 2003), and the lack of *sharing of unique information* so that intellectual resources of a group are underused (Larson et al., 1996, 1998; Stasser, 1999; Wittenbaum & Bowman, 2004). Although this research provides empirical evidence about real people to convincingly test particular hypotheses, it is often limited to small groups with minimal structures performing tightly controlled tasks of limited duration.

In contrast, the approach of multi-agent systems originates in distributed artificial intelligence (Weiss, 1999) and the modeling of complex, adaptive systems, such as animal swarms (Bonabeau, Dorigo & Theraulaz, 1999) and human societies (Epstein & Axtell, 1996; Nowak, Szamrej & Latané, 1990). In these systems, agents interact in order to reach their individual or group objectives, which may be conflicting. However, the combination of their local, individual actions produces emergent behavior at the collective level. In contrast to experimental psychology, this approach has been used to investigate complex situations, such as the emergence of cooperation and culture, and the self-organization of the different norms and institutions that govern the interactions between individuals in an economy. It manages to tackle such complex problems by means of computer simulations in which an unlimited number of software agents interact according to whatever simple or complex protocols the researcher has programmed them to obey. Although the complexity of situations that can be studied is much greater than in an experimental paradigm, there is no guarantee that the results produced by the simulation say anything meaningful about real human interaction.

Moreover, many of the earlier simulations to represent agent interaction are too rigid and simplistic to be psychologically plausible. The behavior of a typical software agent is strictly rule-based: If a particular condition appears in the agent's environment, the agent will respond with a particular, preprogrammed action. More sophisticated cognitive architectures may allow agents to learn, that is, adapt their responses to previous experience, but the learning will typically remain within a symbolic, rule-based level. Perhaps the most crucial limitation of many models is that the individual agents lack their own psychological interpretation and representation of the environment. As Sun (2001, p. 6.) deplored, multi-agent systems need "…better understanding and better models of individual cognition".

Another shortcoming of many multi-agent models is their rigid representation of relations between agents. In the most common models, *cellular automata*, each agent ("automaton") occupies a cell within a geometrical array of cells, typically in the form of a checkerboard. Agents will then interact with all the agents in their geometric neighborhood. This is clearly a very unrealistic representation of the social world, in which individuals interact differentially with other individuals depending on their previous experiences with those others. Other types of simulations, such as *social network* models, are more realistic with respect to human relationships, but still tend to be rigid in the sense that the agents cannot change the strength of their relationship with a particular other agent (for a full discussion, see section on *Alternative Models*).

The present paper presents a first step towards overcoming these various limitations, by proposing an integration of social psychology and multi-agent approaches. For this, we will present a generic multi-agent simulation environment that is as much as possible psychologically plausible, while at the same time remaining as simple as possible. First, each individual agent is represented by a recurrent connectionist network (to be described shortly) that is capable of representing internal beliefs as well as external information. Such a network can learn new associations between its different concepts through experience. This type of model has been used successfully in the past to model several phenomena in social cognition, including person impression formation (Smith & DeCoster, 1998; Van Overwalle & Labiouse, 2004), group impression formation (Kashima, Woolcock & Kashima, 2000; Queller & Smith, 2002; Van Rooy, Van Overwalle, Vanhoomissen, Labiouse & French, 2003), attitude formation (Van Overwalle & Siebler, 2005), causal attribution (Van Overwalle, 1998; Read & Montoya, 1999), and many other social judgments (for a review, see Read & Miller, 1998).

Second, and this is the real innovation proposed in this paper, we extend this connectionist model of individual cognition to the social relations *between* agents. Inspired by previous models developed by Hutchins (1991) and Hutchins and Hazlehurst (1995), our model consists of a collection of individuals networks, that each represent a single individual, and that can communicate with each other. This is depicted in Figure 1 for three agents. In our model, unlike earlier multi-agent models, any agent can in principle interact with any other agent, but the strength of the interaction will adapt to experience. We will introduce the fundamental concept of *cognitive trust* as a measure of the degree to which one agent influences another. Cognitive trust can be defined as the confidence that one agent has in the information communicated by another one. It thus refers to the perceived validity or credibility of received information given the receiver's knowledge about the sender.. Note that this is different from the more common meaning of *trust* which —apart from other specific uses in law and business— can be broadly summarized as the expectation that another agent would perform some action (as part of an exchange, as a result of one's relationship with the other, or of the power one has over the other). In the remainder of the paper, however, we will use the shorthand term *trust* to refer to *cognitive trust*. Communication involves the transmission of information on the same concepts from one agent's network to another agent's network, along connections whose adaptive weights reflect the perceived trust in the information transmitted. Hence, the group of agents functions as an adaptive, socially distributed network where information and knowledge are distributed among and propagated between different individual networks.

To demonstrate that this model is realistic, we will run several simulations that can accurately reproduce several of the key experimental results obtained by psychologists studying communication and cognition in groups. Thus, our model is able to directly integrate the two major approaches towards the study of distributed cognition: social psychology and multi-agent models. We will further situate our model within the broad multi-agent landscape by comparing it with related models, and we end with the implications and limitations of the proposed model and identify areas where further theoretical developments are needed. Suffice it to say here that at present our model is purely focused on the processing and exchange of information: Our agents do not try to reap any goal, reward, or "utility", like in most game-theoretic simulations of economic rationality or cooperation. The only "actions" they perform are the transmission and interpretation of information, based on prior experience. Yet, these simple assumptions will be shown to accurately explain many of the concrete biases that characterize group level cognition. But first, we describe the proposed connectionist model in some detail.

A Connectionist Approach to a Collection of Individual Nets

Given the plethora of alternative models that have been applied in the simulation of collective cognition, one may wonder why a connectionist approach was taken to model individual agents. There are several characteristics that make connectionist approaches very attractive (for an introduction, see McLeod, Plunkett & Rolls, 1998). A first key characteristic is that the connectionist architecture and processing mechanisms are based on analogies with properties of the human brain. Human thinking is seen as an adaptive learning mechanism that develops accurate mental representations of the world. Learning is modeled as a process of on-line adaptation of existing knowledge to novel information provided by the environment. For instance, in group judgments, the network changes the weights of the connections between the target group and its attributes so as to better represent the accumulated history of co-occurrences between the group and its perceived attributes. In contrast, most traditional models in psychology are incapable of such learning. In many algebraic models, recently acquired beliefs or attitudes about target groups or persons are not stored somewhere in memory so that, in principle, they need to be reconstructed from their constituent components (i.e., attributes) every time a judgments is requested or a beliefs is expressed (e.g., Anderson, 1981; Fishbein & Ajzen, 1975). Similarly, activation spreading or constraint satisfaction models recently proposed in psychology can only spread activation along associations but provide no mechanism to update the weights of these associations (Kunda & Thagard, 1996; Read & Miller, 1993; Shultz & Lepper, 1996; Spellman & Holyoak, 1992; Spellman, Ullman & Holyoak, 1993; Thagard, 1989). This lack of a learning mechanism in earlier models is a significant restriction (see also Van Overwalle, 1998).

Second, connectionist models assume that the development of internal representations and the processing of these representations occur in parallel by simple and highly interconnected units, contrary to traditional models where the processing is inherently sequential. The learning algorithms incorporated in connectionist systems do not need a central executive, which eliminates the requirement of centralized and deliberative processing of information. This suggests that much of the information processing within agents is often implicit and automatic. Most often, only the end result of these preconscious processes enters the individual's awareness. Likewise, in human communication, much of the information exchange is outside the agents' awareness, as individuals may not only intentionally express their opinions and beliefs verbally, but they may also unknowingly leak other non-verbal information via the tone of voice, facial expressions and so on (although we do not study these latter aspects).

Finally, connectionist networks have a degree of neurological plausibility that is generally absent in previous algebraic approaches to information integration and storage (e.g., Anderson, 1981; Fishbein & Ajzen, 1975). They provide an insight in lower levels of human mental processes beyond what is immediately perceptible or intuitively plausible, although they go not so deep as to describe real neural functioning. Unlike traditional models, they reveal a number of emergent properties that real human brains also exhibit such as the lack of a clear separation between memory and processing. Connectionist models naturally integrate long-term memory (i.e., connection weights) and short-term memory (i.e., internal activation) with outside information (i.e., external activation). In addition, based on the principle that activation in a network spreads automatically to interconnected units and concepts and so influences their processing, connectionist models exhibit emergent properties such as pattern completion and generalization, which are potentially useful mechanisms for an account of the confirmation bias of stereotype information dissemination within and between agents in a group.

An Individual's Net: The Recurrent Model

An individual agent's processing capacities are modeled by the recurrent auto-associator network developed by McClelland and Rumelhart (1985). We apply this network for two reasons. First, we want to emphasize the theoretical similarities that underlie the present simulations on multi-agent communication with earlier simulation work on social cognition mentioned earlier (e.g., Smith & DeCoster, 1998; Van Rooy et al., 2003; Van Overwalle & Labiouse, 2004; Van Overwalle & Siebler, 2005). Second, this model is capable to reproduce a wider range of social cognitive phenomena and is computationally more powerful than other connectionist models that represent an individual agent's mental processing, like feedforward networks (Van Overwalle & Jordens, 2002; see Read & Montoya, 1999) or constraint satisfaction models (Kunda & Thagard, 1996; Shultz & Lepper, 1996; for a critique see Van Overwalle, 1998).

A recurrent network can be distinguished from other connectionist models on the basis of its architecture (how information is represented in the model), the manner in which information is processed and its learning algorithm (how information is consolidated in the model). It is important to have some grasp of these properties, because the extension to a collection of individual networks described shortly is based on very similar principles.

Architecture

The architecture of an auto-associative network is illustrated in Figure 2 for a talking and a listening agent (the trust weights between agents should be ignored for now and are discussed later). Its most salient property is that all units are interconnected with all the other units (unlike, for instance, feedforward networks where connections exist in only one direction). Thus, all units send out and receive activation. The units in the network represent a target issue (e.g., objects, persons, or groups such as *Jamayans*) as well as various attributes associated with the issue (e.g., behaviors, abilities or traits such a *smart* and *honest*). The connections linking the issue's object with its attributes represent the individual's beliefs and knowledge about the issue. For instance, an individual may believe that waitresses are talkative, and the strength of this belief is represented by the weight of the waitress→talkative connection.

Information Processing

In a recurrent network, processing information takes place in two phases. During the first activation 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. This spreading mechanism reflects the idea that encountering a person, a group or any other object automatically activates its essential characteristics from memory. 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. Typically, activations and weights have lower and upper bounds of approximately –1 and +1.

In non-linear versions of the auto-associator used by several researchers (Smith & DeCoster, 1998; Read & Montoya, 1999), the net activation is determined by a non-linear combination of external and internal inputs updated during a number of internal updating cycles. In the linear version that we use here, the net activation is the sum of the external and internal activations after one updating cycle through the network. Previous simulations by Van Overwalle and colleagues (Van Overwalle et al., 2004, 2005; Van Rooy et al., 2003) revealed that the linear version with a single internal cycle reproduced the observed data at least as well.

Because we will introduce a similar activation updating algorithm for our extension later on, we now present the automatic activation updating in more specific mathematical terms. Every unit *i* in the network receives external activation, termed ext_i, in proportion to an excitation parameter *E* which reflects how much the activation is excited, or

$$a_i = E * ext_i. \tag{1}$$

This activation is spread around in the auto-associative network, so that 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

$$int_i = \underset{j}{\quad} (w_{j\blacklozenge i} * a_j) \tag{2}$$

for all $j \neq i$. The external activation and internal activation are then summed to the net activation, or

$$net_i = E * (ext_i + int_i). \tag{3}$$

According to the linear activation algorithm (McClelland & Rumelhart, 1988, p. 167), the updating of activation at each cycle is governed by the following equation:

$$\Delta a_i = net_i - D * a_i, \tag{4a}$$

where *D* reflects a memory decay term. In the present simulations, we used the parameter values $D = E = 1$ as done previously by Van Overwalle and colleagues. Given these simplifying parameters, the final activation of unit *i* reduces to the sum of the external and internal activation, or:

$$a_i = net_i = ext_i + int_i \tag{4b}$$

Memory Storage

After the first activation phase, 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* (McClelland & Rumelhart, 1988; McLeod, Plunkett & Rolls, 1998).

For instance, if the external activation (e.g., observing a talkative waitress) is underestimated because a weak waitress→talkative connection (e.g., the belief that waitresses are not very talkative) generates a weak internal activation, then the waitress→talkative weight is increased to reduce this discrepancy. Conversely, if the external activation is overestimated because a strong waitress→talkative connection generates an internal activation that is too high (e.g., the belief that waitresses do not stop talking), the weight is decreased. These weight changes allow the network to develop internal representations that accurately approximate the external environment.

In mathematical terms, the delta algorithm strives to match the internal predictions of the network int_i as closely as possible to the actual state of the external environment ext_i, and stores this information in the connection weights. This error reducing process is formally expressed as (see McClelland & Rumelhart, 1988, p. 166):

$$\Delta w_{j\rightarrow i} = \varepsilon * (ext_i - int_i) * a_j, \tag{5}$$

where $\Delta w_{j\rightarrow i}$ is the change in the weight of the connection from unit j to i, and ε 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. When this co-occurrence is not perfect, then the learning algorithm gradually converges to a weight (e.g., +.50) that reflects the proportion of supporting and conflicting co-occurrences (e.g. 50%).

Communication between Individual Nets: the TRUST Extension

The standard recurrent model was augmented with a number of features, which enabled it to realistically reproduce communication between individual agents. This extension assumes that beliefs about objects and their attributes are represented in broadly the same manner among different agents. Communication is then basically seen as transferring the activation on the object and its attributes from *talking* to *listening* agents. This is accomplished by activation spreading between agents in much the same way as activation spreading within the mind of a single individual, with the restriction that activation spreading between individuals is (a) limited to identical attributes and (b) in proportion to the trust connection weights linking the attributes between agents. A crucial aspect of these trust connections is that they reflect how much the information on a given object or attribute expressed by a talking agent is deemed trustworthy, reliable and valid. Thus, the connections through which individuals exchange information are not simply carriers of information, but more crucially, also reflect the degree of trust in this information. This is the cornerstone of our extension to a collection of recurrent networks, and therefore we termed our extended model TRUST.

Because agents can play the role of speaker or listener, the trust connections in the model go in two directions: *Sending* connections for talking and *receiving* connections for listening (see Figure 2) and each agent has both. Note that different trust weights exist for every meaningful object or attribute in the network. Thus, we may trust someone's knowledge on a psychological topic, but distrust his or her sports expertise. Or we may trust someone's ideas on honesty, but disagree that they apply to a particular person or group, or even doubt that we are talking about the same persons after all. Because connectionist systems do not make a principled differentiation between units, all units —be it an individual, social group or their attributes— play the same role in the network. In developing this extension towards multi-agent information exchange, we were strongly inspired by a number of communicative principles put forward by philosophers and psychologists. Specifically, the two trust connections implement to a great deal Grice's (1975) communicative maxims of *quality* and *quantity*.

Maxim of Quality: Sending Trust Weight

The maxim of quality suggests that in order to communicate efficiently, communicators generally try to transmit truthful information ("Try to make your contribution one that is true", p. 46). In the model, the maxim of quality is implemented on the side of the receiving agent. Listeners filter information in proportion to how much they trust that information based on their extant beliefs. This is implemented in *sending* trust connections from the talking to the listening agent for each issue (see Figure 2). When trust concerning a specific issue is maximal (+1), the information expressed by the talking agent is unfiltered by the listening agent. To the degree that trust is lower, information intake by the listener is attenuated in proportion to the trust weight. When trust is minimal (0), no information is processed by the listening agent.

For convenience, we refer to all units involving a communicated issue and its associated attributes as *issue i*. The listener *l* receives information on each issue *i* from all other talking agents *t* in proportion to the trust weights and then sums it. Or, in mathematical terms, for each issue *i*:

$$ext_{\mathrm{li}} = \sum_{\mathrm{t}} (trust_{\mathrm{ti}\to li} * a_{\mathrm{ti}}) \tag{6}$$

where ext_{li} represents the external activation received by the listening agent *l* on issue *i*; $trust_{\mathrm{ti}\to\mathrm{li}}$ is the trust weight from the talking agent *t* to the listening agent *l*; and a_{ti} denotes the activation generated and expressed by talking agent *t*. By comparing with Equation 2, it can be seen that this mechanism of trust spreading between agents is a straightforward copy of activation spreading of standard connectionist models within a single agent.

Maxim of Quantity: Receiving Trust Weights

Grice's (1975) maxim of quantity suggests that communicators transmit only information that is informative and adds to the audience's knowledge ("Make your contribution as informative as is required for the current purpose of the exchange", p. 45). Similarly, research on group minority suggests that communicators tend to increase their interaction with an audience that does not agree with their position. This is implemented in the model by the *receiving* trust weights from the listening agents to the talking agent (see Figure 2). These weights indicate how much the talking agent trusts the listening agent on a particular issue, and hence control how much the taking agent talks about it. To the extent that these trust weights are high (above trust starting weights, $trust_{\circ}$), knowledge and agreement on an issue is assumed and the talking agent will reduce expressing these ideas (attenuation). In contrast, when these weights are low (below trust starting weights, $trust_{\circ}$), the talking agent increases expressing these ideas more strongly (boosting). Since the talking agent may receive different receiving trust weights from different listening agents *l*, we take the maximum[1] receiving trust weight to talking agent *t* on issue *i,* or $trust_{TI} = max(trust_{ti \leftarrow li})$ for all agents *l*. Hence, for each target issue *i* expressed by talking agent *t*:

$$\begin{aligned} &\text{if } trust_{TI} << trust_{\circ} \text{ then } a_{ti} = a_{ti} * (1 + trust_{TI}) \\ &\qquad\qquad\qquad\qquad\quad \text{else } a_{ti} = a_{ti} * (1 - trust_{TI}) \end{aligned} \tag{7a}$$

where a_{ti} is the activation expressed by the talking agent *t* on issue *i* (limited between -1 and +1), and << means that the term on the left is at least a small threshold (of 0.05) smaller than the term on the right. In contrast, for all other issues *j* agent *t* is talking about, the reverse change of activation occurs (limited between -1 and +1):

$$\begin{aligned} &\text{if } trust_{TI} << trust_{\circ} \text{ then } a_{tj} = a_{tj} * (1 - trust_{TI}) \\ &\qquad\qquad\qquad\qquad\quad \text{else } a_{tj} = a_{tj} * (1 + trust_{TI}) \end{aligned} \tag{7b}$$

Note that when an agent talks about several issues, then an interaction ensues between attenuation and boosting. For instance, if listeners agree strongly on two issues *i* and *j*, then issue *i* tends to boost talking about *j* while, conversely, issue *j* tends to attenuate talking about *j* (and similarly for talking about *i*), resulting in little change in the activation of *i* and *j* overall. However, there will be more talking about further issues *k*.

Adjustment of Trust Weights

Given that perceived trust plays a crucial role in the transmission of information, it is important to describe how the trust weights are developed and changed in the model. This adjustment process is alike for receiving and sending trust weights, because they depend on the same underlying process within the listening agent. Like the standard delta learning algorithm where learning is determined by the error between internal and external signal within the individual agents, trust depends on the error between external beliefs (expressed by another talking agent) and the agent's own internal beliefs. For a given issue, if the error within the listening agent is below some *trust tolerance*, the trust weight for that issue is increased towards 1; otherwise, the trust weight is decreased towards 0. In mathematical terms, the change in trust weight by the listening agent *l* on issue *i* expressed by agent *t*, or $\Delta trust_{\mathrm{ti}\rightarrow\mathrm{li}}$, is implemented as follows:

$$\begin{aligned} &\text{if } |\, ext_{\mathrm{li}} - int_{\mathrm{li}} \,| < trust\ tolerance \\ &\text{then } \Delta trust_{\mathrm{ti}\rightarrow\mathrm{li}} = \gamma * (1 - trust_{\mathrm{ti}\rightarrow\mathrm{li}}) * |\, a_{\mathrm{ti}} \,| \\ &\text{else } \Delta trust_{\mathrm{ti}\rightarrow\mathrm{li}} = \gamma * (0 - trust_{\mathrm{ti}\rightarrow\mathrm{li}}) * |\, a_{\mathrm{ti}} \,|, \end{aligned} \tag{8}$$

where ext_{li} represents the external activation received by the listening agent *l* and int_{li} the internal activation generated independently by the listening agent *l*; and γ is the rate by which trust is adjusted and $|\, a_{\mathrm{ti}} \,|$ the absolute value of the activation on issue *i* by the talking agent *t*. In contrast to Equation 5 of the delta algorithm, here the absolute value of the error and the talking activation *i* is taken, because trust is assumed to vary from low (0) to high (+1) only and hence depends on the absolute error between external and internal activation (in proportion to the absolute strength of the activation by which the talking agents express their ideas).

To give an example, imagine that the talking and listening agent agree on the honesty of the Jamayans and both have a connection of +0.80 between these two units (see Figure 2). How does the system, and more in particular the listener, recognize the communication on the attribute *Honesty* as truthful? For the sake of simplicity, we ignore attenuation and boosting (Equation 8). When the talker activates the *Jamayans* unit (with default activation +1), this activation spreads internally to *Honesty* in proportion to the connection weight (+0.80; dotted arrow on the left) and thus becomes +0.80. This activation is then communicated or transmitted along the trust weight (+0.50; dotted arrow on bottom) arriving at the listener with an external activation ext_{li} of +.40 for *Honesty*. Now, something similar happens within the listener. When the talker expresses the *Jamayans* issue, this activation of +1 is transmitted through the trust weight (+0.50) arriving at the listener with an activation of +.50, which is then internally spread within the listener via the connection with *Honesty* (+.080; dotted arrow on the right) resulting in an internal activation int_{li} of +0.40. Thus, the external activation (what the listeners hears) and the internal activation (what the listener beliefs) on the attribute *Honesty* are identical and the talker→listener trust weight involving this attribute will increase. Conversely, as soon as the external and internal activation begin to differ and this difference exceeds the trust threshold, the trust weight will begin to decrease.

Summing up, the larger the sending $trust_{\mathrm{ti}\rightarrow\mathrm{li}}$ becomes, the more the listening agent *l* will trust the talking agent *t* on issue *i* communicated, and the more influential the talking agent will become (maxim of quality). In addition, the larger the receiving $trust_{\mathrm{ti}\leftarrow\mathrm{li}}$ becomes, the more the talking agent *t* will restrain the expression of his or her ideas on this issue (maxim of quantity). A summary of the steps in the simulation of a single trial is given in Table 1.

General Methodology of the Simulations

Having spelled out the assumptions and functioning of the TRUST model, we apply it to a number of classic findings in the literature on persuasion, social influence, interpersonal communication and group decision. For explanatory purposes, most often, we replicated a well-known representative experiment that illustrates a particular phenomenon, although we occasionally also simulated a theoretical prediction. Table 2 lists the topics of the simulations we will report shortly, the relevant empirical study or prediction that we attempted to replicate, as well as the major underlying processing principle responsible for reproducing the data. Although not all relevant data in the vast attitude literature can be addressed in a single paper, we belief that we have included some of the most relevant phenomena in the current literature. However, before discussing these simulations in more detail, we first provide an introductory overview of the different processing phases in the simulations, and we end with the general parameters of the model.

Coding of Learning and Communicating

In all simulations, we assumed that participants brought with them learning experiences taking place before the experiment. This was simulated by inserting a *Prior Learning Phase*, during which we briefly exposed the individual networks to information associating particular issue object with some attributes. Although these prior learning phases were kept simple and short for explanatory reasons (involving only 10 trials), it is evident that real life exposure is more complex, involving direct experiences or observations of similar situations, or indirect experiences through communication or observation of others' experiences. However, our intention was to establish connection weights that are moderate so that later learning still has sufficient impact.

We then simulated specific experiments. This mostly involved a *Talking and Listening Phase* during which agents communicated. As we will see shortly, given that the issue topic was typically imposed externally (by the experimenter), this was designed as external activation and the internal activation spread automatically from this external input reflects the participant’s own thoughts on the issue. Talking was then simulated by spreading these two sources of activation in the talking agent’s units —via the trust weights— to the corresponding listening agents’ units representing the same concepts (see Table 1 for the exact order of the computations). The particular conditions and trial orders of the focused experiments were reproduced as faithfully as possible, although minor changes were introduced to simplify the presentation (e.g., fewer trials or arguments than in the actual experiments).

Measuring Beliefs and Communicated Content

At the end of each simulated experimental condition, test trials were run to assess the dependent variables of the experiments. The specific testing procedures are explained in more detail for each simulation. The obtained test activations of the simulation were then compared with the observed experimental data. We report the correlation coefficient between simulated and empirical data, and also projected the simulated data onto the observed data using linear regression (with intercept and a positive slope) to visually demonstrate the fit of the simulations. The reason is that only the pattern of test activations is of interest, not the exact values.

General Model Parameters

In spite of the fact that the major experiments to be simulated used very different stimulus materials, measures and procedures, all parameters of the auto-associator are set to the same values, unless noted otherwise. As noted earlier, the parameters of the individual nets were the same as in earlier simulations by Van Overwalle and colleagues ($E = D =$ number of internal Cycles = 1, and a linear summation of internal and external activation; see Equation 4). The last parameter implies that activation is propagated to neighboring units and cycled one time through the system. The other parameters are listed in Table 3. In order to ensure that there was sufficient variation in the data to conduct meaningful statistical analyses, all weights were initialized at the specified values plus additional random noise between -.05 and +0.05.

Persuasion and Social Influence

Once people are in a collective setting, it appears that they are only too ready to conform to the majority in the group and to abandon their own personal beliefs and opinions. Although dissenting minorities may possibly also have some impact, the greater influence of the majority is a remarkably robust and universal phenomenon. Two major explanations have been put forward to explain the influence of other people in a group: pressure to conform to the norm and informative influence. This section focuses on this latter informative explanation of social influence: Group members assess the correctness of their beliefs by searching for adequate information or persuasive arguments in favor of one or the other attitude position. Perhaps one of the most surprising upshots of this informative influence or conversion is group polarization —the phenomenon that after a group discussion, members of a group on average shift their opinion toward a more extreme position. The next simulation demonstrates that the number of arguments provided to listeners is crucially important in changing their beliefs. Additionally, we illustrate that the TRUST multi-agent model predicts decreasing group polarization as communication between group factions declines.

Simulation 1: Number of Arguments

Key Experiment. To demonstrate that the sheer number of arguments communicated to other people strongly increases their willingness to adopt the talker's point of view, we now turn to an illustrative experiment by Ebbesen and Bowers (1974, Experiment 3). This experiment demonstrates that shifts in beliefs and opinions are to a great extent due the greater number of arguments received. Participants listened to a tape-recording of a group discussion, which contained a range from little (10 %) to many (90 %) arguments in favor of a more risky choice. As can be seen in Figure 3, irrespective of the arguments heard, the participants shifted their opinion in proportion of the risky arguments heard in the discussion.

Simulation. Table 4 represents a simplified learning history of this experiment. The first five cells reflect the network of the talking agent (i.e., the discussion group) while the next five cells reflect the network of the listener. Each agent has one unit reflecting the issue object (i.e., topic of discussion) and four units to reflect its attributes (i.e., the features of the arguments). Each row reflects a trial or act of learning or communicating, and depending on the specific issues and features present or absent at a trial, the respective units are turned on (activation level > 0) or turned off (activation level = 0). Because this is the first simulation, we describe its progress in somewhat more detail:

- The discussion group from which the recording was taken (talking agent) first learns the features or characteristics involving the discussion topic. This learning was implemented in the first *Prior Learning* phase of the simulation (see Table 4).
- Next, each participant listens to the recorded talks by the discussion group (i.e., talking agent). As can be seen, talking is implemented during the *Talking and Listening* phase by activating the issue object (i.e., topic of discussion) in the talking agent and then allowing the agent's network to spread this around and generate internal activation (i.e., own beliefs). Both sources of activation then initiate the "talking" about one's own opinions, that is, they spread to the listening agent in proportion to the trust weights where it is received (as indicated by a "?"—denoting little ears— in the cells). The varying number of arguments is implemented in the simulation by having 1, 3, 5, 7 or 9 Talking and Listening trials, which corresponds exactly to the number of risky arguments used by Ebbesen and Bowers (1974). In the simulation, we employed the same set of 4 feature units to denote that these units represent essential aspects in all arguments (i.e., that the behavior involves risk), and a repetition of these aspects is what increases the changes in the listener's opinion.
- Finally, after the arguments have been listened to, the listener's opinion is measured in the *Test of Belief* phase by activating (or "priming") the topic in the listening agent, and allowing the neural network of the listener to generate its internal activation (i.e., own beliefs). This internal activation is then recorded (as indicated by "?") and averaged. These simulated data are then projected onto the observed data for visual comparison

Results. The "statements" listed in Table 4 were processed by the network for 50 "participants" with different random orders. In Figure 3, the simulated values (broken lines) are compared with the attitude shifts (striped bars) observed by Ebbesen and Bowers (1974). Because the starting weights of the listener are always 0 (and additional random noise), the simulated data reflect a final attitude as well as an attitude shift. As can be seen, the simulation fits very well and the correlation between simulated and observed data is significant, $r = .98$, $p < .01$. An ANOVA on the simulated attitude further reveals a significant main effect of the proportion of arguments heard, $F(4, 245) = 4.63$, $p < .001$. Note that these results do not change appreciably when attenuation and boosting (maxim of quantity) is turned off in the simulation.

Extensions and Discussion. The simulation demonstrates that some minimal level of trust is needed to pass information from one agent to another. Trust may have several determinants, such as familiarity, friendliness and expertise of the source. Another very powerful determinant of trust is membership of a group. Typically, people trust members from their own group more than members of another group, as dramatically demonstrated by Mackie and Cooper (1984). Using the same paradigm as Ebbesen ad Bower (1974), they found that listeners' attitudes were strongly altered after hearing arguments from an ingroup, but were much less affected when the same arguments came from an outgroup or a collection of unrelated individuals. We successfully simulated this effect by repeating the previous simulation, with the crucial modification that in the ingroup condition, all starting talking→listener trust weights were set to +1 (to reflect that the listeners trust the talking agents completely) and in the outgroup/individual condition, all starting talking→listener trust weight were set to 0 (to reflect complete lack of trust).

Interlude Simulation: Polarization

The TRUST model predicts that persuasive communication alone leads to group polarization because the continuing influence of the majority's opinions gradually shifts the minority's dissident position in majority direction (except, of course, when trust between group members is very low). This prediction is consistent with a meta-analysis by Isenberg (1986) demonstrating that persuasive argumentation has stronger effects on polarization than group pressure or social comparison tendencies to conform to the norm. This prediction is briefly demonstrated in the next simulation. We simulated 11 agents, each having 3 units denoting the topic of discussion, as well as a positive and a negative valence unit denoting the position on the topic (for a similar approach, see Van Overwalle and Siebler, 2005). By providing more or less learning trials with the positive or negative valence units turned on, we manipulated the agents' position on an activation scale ranging between -1 and +1. We then allow all agents to exchange their ideas with all other agents (on each discussion round, each agent talked two times to everybody else). Afterwards, the agents' attitude is measured by priming the topic and reading off the difference between the positive and negative valence units.

As can be seen in the left panel of Figure 4, polarization was already obtained after one discussion round as all agents moved their position in the direction of the majority, and this pattern was further strengthened the more the participants exchanged their ideas. It is important to note that in order to obtain these results, the tolerance parameter was set to a stricter value of 0.10 instead of 0.50. If tolerance was left at the default value of 0.50, all agents in the group converged to the middle position (average activation 0). This suggests that in order to shift the minority to a majority position (in order to obtain polarization), deviance must be tolerated less. This is probably due to the nature of the task. It seems plausible that when people talk about important beliefs and values, they are less likely to change their ideas than when they are informed about an unknown or novel topic on which they have no *a priori* opinion like in Simulation 1.

Previous simulation work with cellular automata (see section on *Alternative Models*) has suggested that deviant minorities are protected against majority influence by the spatial distance between the two groups. We extended this idea by considering *social* distance as a more crucial and realistic parameter, because people may be less influences by others not only because they are farther apart in space, but also in social status, role, group and the like. In line with the dominant effect of persuasion in polarization mentioned earlier (Isenberg (1986), we simulated social distance by restricting the exchange of information between the minority and majority groups (a) to half the amount of the previous simulation or (b) to zero (while keeping the overall communication alike by increasing the communication within groups). As the middle and right panel of Figure 4 demonstrate, under these circumstances the influence of the majority is strongly reduced and even totally abolished.

Communication

Communication is a primary means by which people attempt to influence and convert each other (Kraus & Fussell, 1996, 1991; Ruscher, 1998). Our primary focus is on research exploring information exchange and how this affects participants' beliefs and opinions. These studies go one step further than those from the previous section by studying actual and spontaneous conversation; and by recoding the content of these conversations, data is collected on the natural course of information exchange.

Simulation 2: Referencing

Key Experiment. Several studies explored how people use language to identify abstract or concrete objects. This paradigm is termed *referencing* (e.g., Kingsbury, 1968; Krauss & Weinheimer, 1964, 1966; Schober & Clark, 1989; Steels, 1999; Wilkes-Gibbs & Clark, 1992). Typically, one person is designated as the "director" and is given the task to describe a number of unfamiliar pictures to a "matcher" who cannot see these pictures and has to identify them. In order to provide a satisfactory solution to the task, both participants have to coordinate their actions and linguistic symbols to refer to the pictures and have to establish a joint perspective during the conversation (Schober & Clark, 1989). This collaborative process is also termed *grounding*. The aim of the research is to asses how this perspective-taking and coordination influences the participants' messages and the adequacy of these messages. Figure 5 (top panel) depicts the typical outcome of such studies (Kraus & Fussell, 1991). On their first reference to one of the pictures, most directors use a long description, consisting of pictorial attributes. Next, matchers often ask clarifying questions or provide confirmations that they understood the directions (e.g., Schober & Clark, p. 216). Over the course of successive references, the description is typically shortened to one or two words. Often the referring expression that the conversants settle on is not one that by itself would evoke the picture. This is taken as evidence that people not simply decode and encode messages, but that they collaborate with each other moment by moment to try to ensure that what is said is also understood. Schober and Clark (1989) conducted such a referencing experiment, where a "director" described a number of unfamiliar pictures, and the "matcher" had to identify the intended one from a collection of pictures. Their experiment is interesting, because unlike previous studies, they recorded not only the verbal descriptions of the director, but also the reactions of the matcher (Experiment 1). An analysis of the messages (see Figure 5, bottom panel) shows that both directors and matchers used progressively less words to describe the pictures, although directors —because of their explanatory role in the conversation— used more words than matchers.

Simulation. Table 5 represents a simplified simulation of this experiment. The architecture and learning history are very similar to the previous simulations as it contains two agents, each having one unit to refer to the image and four units to describe its features. For illustrative purposes, we used the features from the Martini example in Figure 5 (top panel). First, the director studied a figure during a *Prior Observation* phase. Next, during the *Talking and Listening* phase, both agents talked and listened to each another in a randomized sequence, with the sole limitation that the director talked more often than the matcher. Although this is clearly far from the content of a natural conversation where a matcher asks for specific clarification and elaboration and the director provides targeted answers and descriptions, as we will see, these minimal assumptions with respect to the amount of talking are sufficient to replicate the basic effect. Of most importance here is that not all features are equally strongly connected with the object. However, rather than allowing the network to randomly generated stronger and weaker connections for some features (which would be more natural as the connection strength differs between the participants), we set the activation values of the two last features to a lower value to make this assumption explicit. To test the content of the conversation, we simply measured the average activation during the *Talking and Listening* phase.

Results. The "statements" listed in Table 5 were processed by the network for 50 "participants" with different random orders. In Figure 5 (bottom panel), the simulated values (broken lines) are compared with the observed number of words (bars). As can be seen, the simulated and observed number of words match very well and the correlation between them is significant, $r = .99$, $p < .001$. An ANOVA on the simulated data further reveals a significant main effect of the number of words for the director, $F(5,294) = 111.01$, $p < .001$, as well as for the matcher, $F(5,294) = 56.26$, $p < .001$.

Extensions and Discussion. The same simulation approach was applied with success on related work on referencing (Kingsbury, 1968; Krauss & Weinheimer, 1964, 1966) with the same parameters, except for a lower initial trust weight in Kingsbury (1968). Schober and Clark (1989) also tested the accuracy of matchers and other people who overheard the communication between director and matchers, and we were also able to replicate these accuracy results.

How does the simulation produce the decreasing pattern of words over time? In the introduction of the TRUST model, we explained that the maxim of quantity is implemented by the trust weights from the listening agent to the talking agent. A high talker←listener weight indicates that the listener is to be trusted, and reduces the expression of topics that the listener is already familiar with, while the expression of other information is boosted. Consistent with the maxim of quantity, when attenuation and boosting was turned off in the present simulation, the pattern of results changed (i.e., the number of words did not decline completely and started to go up again after some time). However, there is also a second, more important reason for the reduced number of words.

Although our network cannot make a distinction between, for instance, descriptions, questions or affirmations, it ensures that only the strongest connections with feature of the picture are reinforced and are ultimately singled out as "pet" words to describe the whole picture. The weaker connections die out because repeating the same information over again "overactivates" the system. Because stronger connections already sufficiently describe the object, the weaker connections are overshadowed by them and become obsolete. This property of the delta algorithm is also known as competition or discounting (see Van Overwalle, 1998; Van Overwalle & Labiouse, 2004; Van Rooy et al., 2003). In the simulation, this overshadowing by stronger connections is clearly demonstrated when all the features are given equal and maximum connection strength of +1 (by providing all an activation of +1) in the Prior Learning Phase. Under this specification, the simulated pattern differs from the observed data (i.e., it showed a less robust and smooth decrease in number of words so that, for instance, the number of word was not always the highest on the first trial).

Thus, whenever people repeat information, our cognitive system make its memory representations more efficient by making reference only the strongest features, while the other features are gradually suppressed. For instance, people typically simplify recurring complexities in the outside environment by transforming them into stereotypical and schematic representations. This points to a cognitive mechanism of efficiency in memory as an additional reason for the decreasing number of words, in addition to coordination and establishment of a joint perspective during the conversation (Schober & Clark, 1989). One might even argue that joint coordination through simplification in the reference paradigm is so easy and natural precisely because it relies on a basic principle of cognitive economy that our brain uses constantly.

Interlude Simulation: Talking Heads

In his "Talking Heads" experiment, Steels (1999, 2003, 2005, Steels & Belpaeme, 2005) describes an open population of cognitive robotic agents who could detect and categorize colored patterns on a board and could express random consonances to depict them. With this *naming game* about real world scenes in front of the agents, Steels (1999) wanted to explore how humans created meaningful symbols and developed languages. Although the present TRUST model obviously does not have the full capacities of Steels' robots, it is able to replicate the naming game that the robots also played. Four different sorts of meaning extraction are potentially problematic in this process: the creation of a new word, the adoption of a word used by another agent, the use of synonyms by two agents, and ambiguity when two agents use the same word for referring to different objects (i.e., homonyms).

To simulate these four types of meaning creation, a talking and listening agent first learned their respective word for two objects, and then after mutual talking and listening (to an equal degree), we tested how the listening agent generates the meaning of a word that

- is *new* for the listener and was used by the talking agent (e.g., for the listening agent a new Circle→"xu" connection is created);
- *matches* the word of the talking agent (for both agents the same Circle→"xu" connection is used);
- is a *synonym* for the same object (for the talking agent Circle→"xu" is used and for listening agent Circle→"fepi");
- is *ambiguous* in that the same word is used for different objects (for the talking agent Circle→"xu" is used and for the listening agent Square→"xu").

Figure 6 displays how the meanings of the words are created and changed under these four circumstances. Note that we turned off the criterion of novelty (attenuation and boosting) so that we could concentrate on the acquisition of meaning rather than the expression and communication of it. As can be seen, the simulated process for the creation of a new word and matching an existing word are obvious. The new word gradually acquires strength and the matched word keeps the high strength it has from the beginning. For synonyms, the simulation reveals a competition between words. To denote a circle, the listening agent gradually loses its preference for "fepi" in favor of the word "xu" that is then used more often (although the synonym "fepi" is still used). For ambiguous words, the simulation predicts no competition so that the ambiguity is not really solved. Instead, "xu" tends toward a meaning at a higher categorical level referring to both objects alike (like the word "geometric figures" refers to circles and squares) although the listening agent keeps its initial preference for Square→"xu" because it is the more prototypical member of the category.

Simulation 3: Stereotypes and the Rumor Paradigm

A surprising finding in recent research on stereotyping is that our stereotypic opinions about others are not only generated by cognitive processes inside people's head (see Van Rooy et al. 2003, for a connectionist view), but are further amplified by the mere communication of these beliefs. Several investigations have demonstrated that information that is consistent with the receiver's beliefs is more readily exchanged, while stereotype disconfirming information tends to die out (Brauer, Judd & Jacquelin, 2001; Klein et al., 2003; Kashima, 2000; Lyons & Kashima, 2003; Ruscher & Duval, 1998; Schulz-Hardt et al., 2000; Thompson, Judd & Park, 2000). This process reinforces extant stereotypes even further, attesting to the crucial role of social communication such as rumors, gossip and so on in building impressions about others.

Key Experiment. The maxim of quality suggests that communication is more effective when the information is considered trustworthy. Hence, we might expect that people with similar stereotypical background (and thus are mutually experienced as trustworthy), exchange their stereotypical beliefs more often. In contrast, people with a different background evaluate each other as less trustworthy. In addition, because they exchange conflicting information (from their opposing background), their messages tend to cancel each other out. To illustrate these predictions of the TRUST model, we simulate a study undertaken by Lyons and Kashima (2003, Experiment 1). This study stands out because the exchange of information was tightly controlled and recorded for each participant. Specifically, information was communicated through a serial chain of 4 people. One person begins to read a set of information before reproducing it from memory to another person. This second person then reads this reproduction before then reporting it verbally to a third person and so on, much the same way as rumors are spread in a community. The information in the study involved a story depicting a member of a fictional group, the "Jamayans". Before disseminating the story along the chain, general stereotypes were induced about this group (the background information). In one of the conditions, all 4 participants in the chain were given the same stereotypes about the Jamayans that they were smart and honest (*actual shared condition*). In another condition, 2 participants were given stereotypes about the Jamayans that were opposite to that given to the other 2 participants, so that each subsequent participant in the chain held opposing group stereotypes (*actual unshared condition*). The target story given afterwards always contained mixed information that both confirmed and disconfirmed the stereotype.

As can be seen in Figure 7 (left panel), when the stereotypes were shared, the reproduction became more stereotypical further along the communication chain. The story was almost stripped of stereotype inconsistent (SI) information, whereas most of the stereotype consistent (SC) information had been retained. In contrast, when the stereotypes were not shared (right panel), the differences between SC and SI story elements were minimal.

Simulation. We simulated a learning history that was basically similar to the original experimental procedures used by Lyons and Kashima (2003, Experiment 1). The architecture involved 5 agents, each having 5 units consisting of the topic of the information exchange (i.e., Jamayans), and two stereotype consistent traits (smart and honest) and two stereotype inconsistent traits (stupid and dishonest). As can be seen in Table 6, for the actual shared condition, we provided during the *Prior SC Information* phase, 10 stereotypical trials indicating that the Jamayans were smart (by activating the Jamayans and the smart unit) and 10 stereotypical trials indicating that they were honest (by activating the Jamayans and the honest unit) for each of the agents. For the actual unshared condition, agents 2 and 4 received contradictory information indicating that the Jamayans were stupid and dishonest (by activating the stupid and liar units) during the *Prior SI Information* phase. Next, during the *Story* phase, the first agent received ambiguous information about a member of the Jamayans: 5 SC trials reflecting story elements indicating that he was smart and 5 SI story elements indicating that he was a liar. This story was then reproduced by this agent and received by the next agent. That is, the Jamayans unit in agent 1 was activated and, together with the internal activation (i.e., expression of beliefs) of the other smart/stupid and honest/liar units in agent 1, was then transmitted to agent 2. After listening, agent 2 expressed his or her opinion about the Jamayans to agent 3, then agent 3 to agent 4, and finally agent 4 to agent 5 (the participants in the fourth position of the experimental chain were led to believe that there actually was a fifth participant). After each Talking and Listening phase, we measured how much the talking agent had expressed or communicated the notion that the Jamayans were smart, stupid, honest or liar

Results. We ran the "statements" from Table 6 for 50 "participants" in the network with different random orders. As can be seen in Figure 7, the simulation closely matched the observed data (r = .93, p < .001). Separate ANOVAs with Sharing (shared vs. unshared) and Position (1, 2, 3 or 4) as factors, reveal a significant interaction for both the SC and SI story element, F(3, 392) = 19.29—19.93, p < .001. Separate t-tests show that all adjacent positions differed reliably; except for SC information in the shared condition. This suggests that in the shared condition, the amount of SC information transmitted from one agent to the other was almost completely maintained in contrast to the SI information which decreased. In the unshared condition, the expression of both types of information decreased. Note that boosting or attenuation of belief expression (maxim of quantity) should not play a role here as the agents had no opportunity to hear their communication partners before telling their own story, and so were unable to test whether they agreed on the Jamayans' attributes. To verify this, we ran the simulation with boosting and attenuation turned off, and found identical results. This strongly suggests that for a rumor paradigm involving novel interlocutors, the initial trust in a talking agent's statements (maxim of quality) was sufficient for creating a stereotype confirmation bias during group communication.

Extensions. We also successfully simulated related work on stereotyping and impression formation using the rumor paradigm and free discussions (but without recording and analysis the conversation itself), such as Thompson, Judd and Park (2000, Experiments 1 & 2), Brauer, Judd and Jacquelin (2001, Experiment 1), Schultz-Hardt, Frey, Lüthgens & Moscovici (2000, Experiment 1) and Ruscher & Duval (1998, Experiment 1). Although we had to change our parameters somewhat in some of the simulations (e.g., changing the initial trust or tolerance levels), overall, this attests to the wide applicability of our approach.

Simulation 4: Maxim of Quantity and the Expression of Stereotypes

The maxim of quantity suggests that when the audience is knowledgeable and agrees with the communicator's position, less information is transmitted. Recall that in the model, the maxim of quantity (implemented by a strong talker←listener trust weight) indicates that the listener is to be trusted and that expression of the same issue or attribute can be attenuated, while the expression of other issue or attribute should be boosted.

Key Experiment. To illustrate the working of the maxim of quantity, we now apply the TRUST model to another data set from the same empirical study by Lyons and Kashima (2003) described above. In this study, Lyons and Kashima provided half of their participants with the false information that the other participants in the chain had received completely similar background information on the Jamayans (*perceived complete knowledge*), while the other half were given the false information that the other participants were completely ignorant (*perceived complete ignorance*). As shown in Figure 8, the results indicated that given the belief of complete knowledge, both SC and SI story elements were reproduced and no substantial stereotype bias emerged. In contrast, in the complete ignorance condition, a stereotype bias became apparent in that SI story elements were strongly suppressed.

Simulation. We ran the same simulation as before, with the following modifications. In order to obtain high trust weights from the listening agents to the talking agents, (a) we included only the actual shared condition, and (b) in the *perceived complete knowledge* condition, we set the initial trust weights from listening to talking agents 0.20 above the trust starting weight for the units involved in the transmission of SC information (Jamayans, smart, honest). These high trust weights directly simulate the notion that the listening agents were to be trusted more than usual because they share the same background with the speaker.

Results. As can be seen in Figure 8, the simulation matched the observed data although not above conventional levels of significance (r = .81, p = .19) partially due to lack of data points (only 4), and partly because of the implementation of the maxim of quantity. If only boosting on other issues were implemented (without attenuation of familiar issues) then the simulation would match almost perfectly the observed data. However, because attenuation of known information is the core idea of the maxim of quantity (see Grice, 1975), we left this crucial aspect intact. An ANOVA revealed the predicted significant interaction between perceived Sharing (knowledge vs. ignorance) and Type of Information (SC versus SI), F(1, 796) = 139.92, p < .001. Further t-tests revealed that although the difference between SC and SI was still significant under complete knowledge, t(398) = 4.45, p < .001, it was much less so than under complete ignorance, t(398) = 23.25, p < .001. The implementation of the maxim of quantity was crucial in the simulation of higher SI information in complete knowledge condition. This strongly suggests that when people believe they share a similar background, the maxim of quantity helps to neutralize the stereotype confirmation bias in communication.

Simulation 5: Group Problem-Solving and Sharing Information

It is often believed that a group as a whole has more intellectual resources to help solve a problem than individuals, because some members may have crucial information that others do not have and that may lead to a solution. By pooling all such unique information, the group as a whole should make considerable better decisions. However, contrary to this ideal of group problem solving, research has revealed that unique information is discussed less often than shared information, and if it is, it is often brought in the discussion much later (Stasser, 1999; Stasser and Titus, 1985). This, of course, reduces the efficiency and speed of group problem solving (Larson, Christensen, Abbott & Franz, 1996; Larson, Foster-Fishman & Franz, 1998).

This result is unexpected. Would one not expect on the basis of Grice’s (1975) maxim of quantity, that group members discuss known or shared information less in favor of novel information? Why, then, are they doing the opposite? One explanation, put forward by Stasser (1999) is that shared information has a sampling advantage. That is, because shared information has a higher probability of being mentioned than unique information (since many members hold shared information), groups tend to discuss more of their shared than their unshared information. However, each time an item of shared information is brought forth, because most group members hold this just-mentioned item, the probability to sample additional shared information is reduced more than unique information. This sampling explanation predicts more discussion of shared information at the start of a discussion, while unique information is brought in the discussion later. Another explanation, based on the idea of *grounding*, was put forward by Wittenbaum and Bowman (2004). They argued that group members attempt to validate one's thoughts and ideas by assessing their accuracy and appropriateness through comparison with others. Shared information can be socially validated and hence provides opportunities to evaluate each other's task contributions, while unshared information cannot. Therefore, it is exchanged more often.

The idea that people validate their thoughts and ideas through comparison with others is also at the core of our approach. Indeed, social validation and trust depend on the information being consistent with one's beliefs. As put forward by Stasser (1999), because of uneven sampling probabilities, shared information is communicated more often at the beginning of a group discussion. However, after the information has been validated, trust is high and the maxim of quantity kicks in. This implies that after a while, discussion of shared information is attenuated while discussion on other issues is boosted, and so increases the discussion of unique ideas.

Key Experiment. To illustrate the working of the maxim of quantity in group problem solving, we conduct a simulation of an empirical study by Larson et al. (1996). Participants in this study watched a short videotaped interview of a patient by a physician in an examination room. Three different videotapes were created. Roughly half of the information in each tape was also present in the other tapes (shared condition) while the remaining information was present in one tape alone (unique condition). Each videotape was shown to different teams consisting of a mixture of medical students and established experts. Afterwards, each team discussed the case and produced a differential diagnosis. This discussion typically lasted less than 20-25 minutes. Figure 9 shows the percentage of shared as opposed to unique information as the discussion unfolded, where each *discussion position* reflects the point in the discussion where a new item was introduced (i.e., omitting the repetitions). As can be seen, there is a negative linear trend in that initially a lot of shared items is discussed, while at the end more unique items are mentioned.

Simulation. We simulated a discussion by 3 agents (although only 2 are shown in Table 7). Each agent had 7 units (although only 5 are shown) consisting of the object of discussion (i.e., the patient), and 3 shared items and 3 unique items. To simulate the viewing of the videotape, during the *Learning* phase, we ran 5 trials for each agent, in which all the shared items were learned and a single unique item. Next, in the *Talking and Listening* phase, we let all agents freely talk about all the shared and unique items. Because agents knew only about a single (unique) item at the beginning of the discussion, this actually provides a sampling advantage for the shared information at the start. To avoid an *a priori* sampling advantage in the simulation, we used the minimal assumption that each agent expressed a single shared and a single unique item during each discussion round. Given the interplay between attenuation and boosting, this leaves each agent in the simulation completely free to talk as much as he or she "wants" about the issue. After each round, we measured how much each agent had communicated the shared and unique items.

Results. We ran the "statements" from Table 7 for 50 "participants" with different random orders. As can be seen in Figure 9, the simulation closely matched the observed data (r = .85, p < .001). A one-way ANOVA on the percentage shared information reveals the predicted main effect of the discussion position, F(34, 2665) = 335.98, p < .001. This confirms that according to the simulation, as the discussion progresses, more unique information is transmitted.

Extension. Research by Stewart and Stasser (1995) indicates that when members are explicitly told about their relative expertise (i.e., their knowledge of their unique information), then the communication of unique information is facilitated. To simulate this effect, we set all receiving trust weights to +1 for the unique information on which the agent was the sole expert, and then ran the simulation again. As can be seen in the figure, more unique information is communicated under these conditions, consistent with the empirical findings. Note that providing high trust weights to all information, including shared information, does not lead to this effect since that gives again a sampling advantage to shared information.

We also simulated a well-known study by Stasser and Titus (1985) that illustrates the lack of sharing unique information. This study served as input for the DISCUSS computer simulation developed by Stasser (1988) to explore and illustrate his theoretical ideas about the integration of information in group discussion and decision making. We did not select this study for the present paper, in part because the data input and simulation is somewhat more complex and because it did not provide data on actual information exchange, but only the end result in participants' beliefs. Nevertheless, a simulation with our model and the same parameters (except the learning rate which was reduced to .25 for more robust results) replicated the major significant results and yielded a mean correlation of r = .79 with observed post-discussion preferences (Stasser and Titus, 1985, Table 5).

Comparisons with Other Models

It is not the first time that computer modeling has been used to aid our understanding of social relationships and influence and to verify the assumed theoretical processes underlying these phenomena. Several categories of computer models can be distinguished: cellular automata, social networks and different types of neural networks (see also Nowak, Vallacher & Burstein, 1998). We describe each of these approaches and discuss their shortcoming and strengths, and compare them to the present model.

Cellular Automata and Social Networks

In cellular automata and social networks, the units of the model represent single individuals and the connections the relationships between individuals. This type of models deals less with processes *within* an individual, but rather *between* individuals.

Cellular Automata. Cellular automata consist of a number of agents (automata) arranged in a regular spatial pattern such as a checkerboard. Each automaton possesses a limited number of attributes (typically 1 or 2) that can be in a limited (often binary) number of states, such as cooperate or defect, or positive or negative opinion. Typically, individuals are linked to their neighbors, and the nature of the linkage is changed by an updating algorithm. Depending on the specific models, these links may represent different characteristics such as persuasiveness, propensity for cooperation, decision strategies and so on (Nowak et al., 1998). Cellular automata allow capturing of regularities and patterns of social norms or opinions in a group while individual agents update their behavior solely on the basis of local information between neighboring agents. For instance, the pioneering work by Nowak, Szamrej and Latané (1990) on the *social impact model* gave rise to important insights such as the role of geographical clustering in protecting deviant minority opinions from being overthrown by a vast majority (see also Interlude Simulation ii). Likewise, Axelrod's *culture* model (1997; see also Axelrod, Riolo & Cohen, 2002; Riolo, Cohen & Axelrod, 2000) showed that a whole population can converge to a local common culture when agents adopt their attributes on the basis of their similarity with neighboring agents (producing roughly an effect like trust weights). In more recent work, Barr (2004) demonstrated convergence to spatially organized symbolic systems (such as a common language or several dialects) and Couzin, Krause, Franks and Levin (2005) illustrated that a small proportion of informed individuals can guide a group of naïve individuals towards a goal. Kennedy (1999) proposed a more sophisticated *particle swarm* algorithm that implements error-driven learning from an agent's own experiences or other neighboring agents, and where agents conform to their neighbors not on the basis of a single attribute like similarity as in previous models, but when their (learned) performance is better than their own. Although cellular automata are flexible with respect to the types of connections they support, they are very rigid with respect to the structure of the social relationships, in that individuals are only influenced by their close neighbors and less (or negligible) so when they are farther apart in space. Hence, social relations are strongly determined by the geometry of the social space, rather than being based on individual choices (Nowak et al., 1998).

Social Networks. This geometrical limitation is relaxed in social networks, where social relations between individuals are depicted as connections among units in a graph. This makes it possible to describe social properties of individual agents (e.g., popularity versus isolation) as well as properties of the groups of individuals (e.g., clustering of opinions in cliques). For instance, Janssen and Jager (2001) developed a *consumat* approach where agents decide to consume products that have the highest personal preference or social preference (e.g., have the largest market shares). However, a limitation of social networks is that the links are often specified in binary positive-negative (or present-absent) terms and do not allow for graded strength. Perhaps more importantly, these models do not provide a general theoretical framework to update the connections in the network. Rodriguez and Steinbock (2004) recently developed a social network model that included graded and adjustable trust relationships between individuals, which appear to better capture the representativeness of decision outcomes. However, given the lack of a general framework, the proposed solution in this as well as in other social networks is often idiosyncratic.

Attractor or Constraint Satisfaction Networks

Like cellular automata and social network, the units of attractor or constraint satisfaction networks represent single individuals and the connections the relationships between individuals. However, unlike these previous models, the connections have graded levels of strength and their adjustments are driven by general algorithms of weight updating. Specifically, the computations for spreading of information and updating relationships are adopted directly from neural networks. Typically, the architecture is a recurrent structure (like the present model), so that all individuals have unidirectional connections with all other individuals.

Many attractor models represent opinion change as the spreading of information or beliefs across individuals. This social spreading is formalized in a similar manner as the spreading of activation in neural models. Eventually, after going through multiple cycles, the model reaches an equilibrium in which the state of the units do not change any more, that is, the model settles in a stable attractor that satisfies multiple simultaneous constraints represented by the supporting and conflicting states and connections of the other units. Hence, an attractor reflects a stable social structure in which balanced social relationships or attitudes are attained. Nevertheless, attractor networks have a fundamental limitation as models of social relations. Since they are based on an analogy with the brain, Nowak et al. (1998) warned that

> it is important to remember that neurons are not people and that brains are not groups or societies. One should thus be mindful of the possible crucial differences between neural and social networks… [some] differences … may reflect human psychology and are difficult to model within existing neural network models (p. 117-118).

Perhaps the most crucial limitation in this respect is that the individuals in the networks do not have a psychological representation of their environment so that individual beliefs are reduced to a single state of a single unit. This shortcoming was overcome by Hutchins (1991). He used constraint satisfaction networks to capture an individual's psychology and memory. Thus, an individual's belief was viewed as the formation of a coherent interpretation of elements that support or exclude each other, in line with earlier constraint satisfaction models of social cognition (Kunda & Thagard, 1996; Read & Miller, 1993; Shultz & Lepper, 1996; Spellman & Holyoak, 1992; Spellman, Ullman & Holyoak, 1993). Crucially, he combined these individual constraint satisfaction networks into a community of networks so that they could exchange their individual information with each other. One of the parameters in his simulations is the persuasiveness of the communication among individuals' nets, which is related to our trust weights although it was only incorporated as a general parameter. A similar approach for two individual nets was developed by Shoda, LeeTiernan and Mischel (2002) to describe the emergence of stable personality attractors after a dyadic interaction. However, in this model, information exchange was accomplished by interlocking the two nets directly (as if two brains were interconnected without any check or control on the veracity or credibility of the information coming from another person), which is implausible as a model of human communication.

Assemblies of Artificial Neural Networks

The constraint satisfaction model of Hutchins (1991) discussed in the previous section was a first attempt to give individuals their own psychological representation and memory. However, one major shortcoming is that the connections in the individual constraint satisfaction networks are not adaptive, that is, they cannot change on the basis of previous experiences (see also Van Overwalle, 1998). This limitation was addressed by Hutchins and Hazlehurst (1995) in a model that consists of a collection of recurrent nets. Each individual was represented by a single recurrent net with adaptive weights (and with hidden layers), and all the nets could exchange information with each other. The model illustrated how a common lexicon is created by sharing information between individuals. However, as we understand it, the output of one individual's network served as direct input for another individual's network as if two brains were directly interlocked with each other, without moderation of the perceived persuasiveness or validity of the received information. This limitation was overcome in the present TRUST model. The *Clarion* model developed by Sun and Peterson (1999) combines information encoding and learning at the symbolic level as well as the lower connectionist level, and permits that agents influence each other at both levels. Possibly, to date, this is the most sophisticated model in this approach, apart from the robots developed by Steels (1999, 2003; Steels & Belpaeme, 2005) which give the individual networks a body with primitive perception and movement.

What has been Accomplished? Limitations and Implications

The TRUST model does more justice to the complexity of the human mind in comparison with earlier multi-agent models, by zooming in on the individual level. This allowed simulating various aspects of human influence and communication in smaller groups as uncovered in empirical research. This constitutes important "post"dictions of our TRUST model. Other simulations also demonstrate that scaling up the TRUST model to larger group phenomenon such as polarization was successful, and suggests that the trust model may be applied to societal phenomena at large (although this is beyond the scope of this article).

Robustness of the Simulations

To what extent are the TRUST simulations dependent on the specifics of the current architecture or parameters? Earlier multi-agent simulation work was often limited in the range of phenomena simulated or used a slightly different model for each. Admittedly, for some of the extensions and interlude simulations, we occasionally set the parameters to values other than the default to replicate a particular study or domain of investigation. Perhaps, this is not surprising given that we explored such different contexts. However, it is important to establish the robustness for the major simulations 1—5 reported earlier, and to do this, we ran these simulations again with different parameter ranges and a different architecture.

First, we performed an exhaustive search of the novel trust parameters (i.e., trust learning rate, trust tolerance and trust starting weights, see also Table 3) in a range between 0.30 and 0.60 (or ± 0.10 all default trust parameters) and identified the lowest correlation between simulated and observed data. For all major simulations except one, the lowest correlation was generally only .02 lower than the correlation with default parameters reported earlier, and the simulation showed the same substantive pattern as the human data. The only exception was Simulation 4 on perceived sharedness (Lyons and Kashima, 2003; lowest $r = .60$) which sometimes underestimated the proportion of stereotypical utterances expressed in the shared condition. Although this result may reflect plausible dependencies on trust parameters that exist also in the real world (such as the degree of knowledge by others), as far as we know, Lyons and Kashima's manipulation has not yet been replicated so that the implications are unclear.

Second, for ease of presentation of the major simulations, up till now we used a *localist* representation where each unit represents a single symbolic concept. However, a more realistic code is a *distributed* representation where each concept is represented by a pattern of activation across a set of units that each represents some subsymbolic micro-feature of the concept (Thorpe, 1994). We reran the major simulations with such a distributed architecture[2] and the results showed on average the same correlation with the human data as the major simulations reported earlier. In sum, then, the major simulation results hold across a wide range of parameter settings and stimulus distributions.[3]

Limitations

However, it is also evident that we have charted only the first modest steps in modeling human collective intelligence through communication. Several aspects have been left out in order to focus on the most important social aspects of communication in a small group of individuals.

First, our model leaves unspecified by what linguistic means information is encoded and decoded during transmission from speaker to listener. However, at least at the moment, the omission of language as a tool of communication seems a sensible simplification which already leads to interesting simulation results that replicate existing research data. Some theorists attempted to develop connectionist models in which language is seen as an essential means of human communication that emerges from the demands of the communication rather than from anything inside the individual agents (Hutchins & Hazlehurst, 1995; Hazlehurst & Hutchins, 1998; Steels, 1999; Steels & Belpaeme, 2005).

Second, our model also largely ignores intentional acts by the communicators. To name a few, it does not incorporate communicative acts such as demands, promises or other requests, or other strategic decisions to fit the information better to the audience's knowledge and background. It also leaves out more subtle strategic maneuvering to please the audience so as when speakers believe that they can make a better impression by speaking with praise about a person that is liked by the audience (Echterhoff, Higgins & Groll, 2005; McCann & Higgins, 1992; McCann, Higgins & Fondacaro, 1991; Schaller & Conway III, 1999). Because such strategic acts are often made deliberatively, they are beyond the current connectionist approach.

Third, a question left unanswered is how trust about agents themselves, such as experts, can be built in the model. Most previous multi-agent systems are incapable of developing and applying such knowledge about trustworthy agents. Although space limitations prevented us to elaborate on this in detail, let us briefly note that the present model can incorporate the meta-notion about an "expert agent" as an additional unit in an individual network, so that trust on a given topic carries over via internal connections to a potential expert source. In their single-agent recurrent connectionist model, Van Overwalle and Siebler (2005) furthermore demonstrated how knowledge about source expertise influences one's opinion on an issue.

What to do Next? Novel Hypotheses

Although incorporating earlier theories and data in a single model is an accomplishment in its own right, theory building is often judged by its ability to inspire novel and testable predictions. In this section, we focus on a few novel predictions involving cognitive trust in information, because this is a novel contribution of the TRUST model that we see as theoretically most crucial in a communicative context. Although there is increasing awareness that trust is an important "core motive" in human interaction (Fiske, 2005), there has been little empirical studies on cognitive trust in social cognition, let alone on its specific role in human communication and information exchange.

What are the determinants and consequences of cognitive trust? Our model proposes that if no a priori expectations on agents exist, people's trust of information is determined by the *fit with their own beliefs*. Although some degree of divergence is tolerated, if the discrepancy is too high, the information is not trusted. Thus, rather than some internal inconsistency or some internal ambiguities in the story told, it is the inconsistency with one's own beliefs that compels the listener to distrust the information. With respect to the consequences of cognitive trust, our model predicts that more cognitive trust results in more and easier adoption of the expressed views and solutions. In contrast, it takes more time to read and judge untrustworthy messages because as they conflict with one's own beliefs, they require scrutiny of multiple interpretations or more elaboration to be accepted or rejected.

Is cognitive trust applied automatically, outside of consciousness? Because acceptance of information on the basis of trust is based on similar processes in the TRUST model as activation spreading in standard connectionist models, we suggest that this is a quite automatic process. Likewise, because the change of trust weights in the model is a straightforward extension of error-driven learning algorithms in connectionist models, we suggest that cognitive trust is changed automatically. Consequently, we predict that inferences on cognitive trust are developed spontaneously on the face of information given, and that social inferences and decisions are made only when the information is (automatically) seen as trustworthy. There is some initial support for these predictions, as Schul, Mayo and Burstein (2004) recently found that trusted messages spontaneously elicit congruent associations while distrust leads to the spontaneous activation of associations that are incongruent with a suspected message. In contrast, although automatic to some degree, we expect that other criteria such as novelty and attenuation of talking about known information can be more easily overruled by controlled processes, such as task instructions and goals, since the act of speaking itself is largely within the control of the individual.

An a more societal level, another hypothesis worth further investigation is suggested by our simulation on the diminished adoption of a majority position when the communication between minority and majority groups is decreased, leading to more divergent opinions (see Figure 4). The interesting question now is: What is the exact amount and type of communication needed for two or more groups to either merge their opinions or remain separate? This problem has immense potential applications, since it describes the all too common situation in society where two cultures or subcultures live in close contact. In some cases the minority is assimilated in the majority (e.g., most initial European communities in the U.S.), in other cases there is strong and apparently growing polarization between groups (e.g., Islamic communities in Europe), leading to potentially violent outbursts. Our model suggests at least two factors that influence the outcome: (a) the amount of communication outside the community because the more people communicate outside their group the smaller the probability of divergence between groups; (b) the degree of initial inconsistency between the opinions in the two communities (e.g. different beliefs on wearing head scarves by women which is obligatory according to Islamic extremists and unwanted by a European majority). Further simulations with more agents, more differentiated beliefs, and a more fine-grained manipulation of the amount of communication between the groups should throw more light on the precise conditions in which two cultures will either merge or divide.

Conclusion

The proposed multi-agent TRUST connectionist model combines all elements of a standard recurrent model of impression formation that incorporates processes of information uptake, integration and memorization, with additional elements reflecting communication between individuals. Although one of many possible implementations, our model reflects the theoretical notion that information processing within and between individuals can be characterized by analogous, yet unique processing principles.

One of the unique features of the model is the assumption that acquired cognitive trust in the information provided by communicators is an essential social and psychological requirement of communication. This was implemented through the inclusion of trust weights, which change depending on the consistency of the incoming information with the receiving agents' existing beliefs and past experiences. Trust weights lead to a selective filtering out of less reliable data and selective propagation of novel information, and so bias information transmission. From this implementation of cognitive trust emerged Grice's (1975) maxims of quality and quantity in human communication. In particular, the maxim of quality was implemented by outgoing trust weights which led to an increased acceptance of stereotypical ideas when communicators share similar backgrounds, while the maxim of quantity was simulated by attenuation in the expression of familiar beliefs (as determined by receiving trust weights) which led to a gradual decreased transmission of stereotypical utterances. By spanning a diversity of findings, the present simulations demonstrate the broad applicability and integrative character of our approach. This may lead to novel insights for well-known social-psychological phenomena and may point to potential theoretical similarities in less familiar domains. It may help us to understand why communicating information among the members of a group sometimes makes their collective cognition and judgments less reliable.

One element that distinguishes the present model from earlier approaches, but that is common to most connectionist modeling, is the dynamic nature of our system. It conceives communication as a coordinated process that transforms the beliefs of the agents as they communicate. Through these belief changes it has a memory of the social history of the interacting agents. Thus, communication is at the same time a simple transmission of information about the internal state of the talking agent, as well as a coordination of existing opinions and emergence of novel beliefs on which the conversants converge and so lead to a "common ground". By combining all those elements in a social-distributed network of individual networks, the unique contribution of our model is that it extends distributed processing, which is sometimes seen as a defining characteristic of connectionism, into the social dimension. This is how social psychology can contribute to the further development of connectionism as tool of theory construction (cf. Smith, 1996).

References

Anderson, N. H. (1981). Foundations of information integration theory. New York: Academic Press.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution, 41,* 203-226.

Axelrod, R., Riolo, R.L., Cohen, M.D. (2002). Beyond geography: Cooperation with persistent links in the absence of clustered neighborhoods. *Personality and social psychology review, 6,* 341-346.

Barr, D. J. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science, 28*, 937—962.

Bonabeau E., Dorigo M. and Theraulaz G. (1999) Swarm intelligence: From natural to artificial systems. Oxford University Press: Oxford, UK.

Brauer, M., Judd, C. M., & Jacquelin (2001). The communication of social stereotypes: The effects of group discussion and information distribution on stereotypic appraisals. *Journal of Personality and Social Psychology, 81*, 463—475.

Couzin, I. D., Krause, J., Frank, N. R., & Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature, 433*, 513—516.

Ebbesen, E. B. & Bowers, R. J. (1974). Proportion of risky to conservative arguments in a group discussion and choice shift. *Journal of Personality and Social Psychology, 29,* 316—327.

Echterhoff, G., Higgins, E. T. & Groll, S. (2005). Audience-tuning effects on memory: The role of shared reality. *Journal of Personality and Social Psychology,* in press.

Epstein J.M. & R. Axtell (1996): Growing Artificial Societies: Social Science from the Bottom Up. MIT Press: Cambridge, MA.

Fishbein, M., & Ajzen, I. (1975). *Belief attitude, intention and behavior an introduction to theory and research*. London, UK: Addison-Wesley.

Fiske, S. F. (2005). Social beings: A core motives approach to social psychology. Hoboken, JN: Wiley.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and semantics: Speech acts (pp. 41—58). New York: Academic Press.

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.

Heylighen F. (1999). Collective Intelligence and its Implementation on the Web: Algorithms to develop a collective mental map. *Computational and Mathematical Organization Theory, 5*, 253-280.

Hutchins, E (1995). *Cognition in the Wild.* MIT Press.

Hutchins, E. & Hazlehurst, B. (1995) How to invent a lexicon: The development of shared symbols in interaction. In G. N. Gilbert & R. Conte (Eds.) *Artificial societies: The computer simulation of social lif*e (pp. 157—189) London, UK: UCL Press.

Hutchins, E. (1991). The social organization of distributed cognition. In L. Resnick, J. Levine, S. Teasley (Eds.) *Perspectives on socially shared cognition* (pp. 283—307). Washington, DC: The American Psychological Association.

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology, 50*, 1141—1151.

Janssen, M.A., & Jager, W. (2001). Fashions, habits and changing preferences: Simulation of psychological factors affecting market dynamics. *Journal of economic psychology, 22,* 745-772.

Kashima, Y. (2000). Maintaining cultural stereotypes in the serial reproduction of narratives. *Personality and Social Psychology Bulletin, 26*, 594—604.

Kashima, Y., Woolcock, J., & Kashima, E. S. (2000). Group impression as dynamic configurations: The tensor product model of group impression formation and change. *Psychological Review, 107*, 914—942.

Kennedy, J. (1999). Minds and cultures : Particle swarm implications for beings in sociocognitive space. *Adaptive behavior, 7,* 269-288.

Kingsbury, D. (1968). *Manipulating the amount of information obtained from a person giving directions*. Unpublished honors thesis, Harvard University, Cambridge, MA

Klein, O. Jacobs, A., Gemoets, S. Licata, L. & Lambert, S. (2003). Hidden profiles and the consensualization of social stereotypes: how information distribution affects stereotype content and sharedness. European Journal of Social Psychology, 33, 755—777.

Krauss, R. M. & Fussell, S. R. (1991). Constructing shared communicative environments. In L. Resnick, J. Levine, S. Teasley (Eds.) *Perspectives on socially shared cognition* (pp. 172—199). Washington, DC: The American Psychological Association.

Krauss, R. M. & Fussell, S. R. (1996). Social psychological models of interpersonal communication. In T. Higgins & A. W. Kruglanski (Eds.). Social psychology: Handbook of basic principles (pp. 655—701). New York, NY: Guilford Press.

Krauss, R. M. & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science, 1*, 113—114.

Krauss, R. M. & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343—346.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory**.** *Psychological Review, 103*, 284-308

Larson, Jr., J. R., Christensen, C., Abbott, A. S. & Franz, T. M. (1996). Diagnosing groups: Charting the flow of information in medical decision-making teams. *Journal of Personality and Social Psychology, 71*, 315—330.

Larson, Jr., J. R., Foster-Fishman, P. G. & Franz, T. M. (1998). Leadership style and the discussion of shared and unshared information in decision-making groups. *Personality and Social Psychology Bulletin, 24*, 482—495.

Lévy P. (1997). *Collective Intelligence*. Plenum.

Lyons, A. & Kashima, Y. (2003) How Are Stereotypes Maintained Through Communication? The Influence of Stereotype Sharedness. *Journal of Personality and Social Psychology, 85*, 989-1005.

MacDonald, T. K. & Zanna, M. P. (1998). Cross-Validation Ambivalence toward social groups: Can ambivalence affect intentions to hire feminists? *Personality and Social Psychology Bulletin, 24*, 427—441.

Mackie, D. & Cooper, J. (1984) Attitude polarization: Effects of group membership. *Journal of Personality and Social Psychology, 46 (3),* 575—585.

McCann, C. D. & Higgins, T. E. (1992). Personal and contextual factors in communication: A review of the 'Communication Game'. In G. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 144—172). London, UK: Sage.

McCann, C. D. Higgins T. E. & Fondacaro, R. A. (1991). Primacy and recency in communication and self-persuasion: how successive audiences and multiple encodings influence subsequent evaluative judgments. *Social Cognition, 9*, 47—66.

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

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

McLeod, P., Plunkett, K. & Rolls, E. T. (1998). *Introduction to connectionist modeling of cognitive processes.* Oxford, UK: Oxford University Press.

Nowak, A. Vallacher, R. R., & Burnstein, E. (1998). Computational social psychology: A neural network approach to interpersonal dynamics. In Liebrand, W. B. G., Nowak, A., & Hegselmann, R. (Eds.) *Computer modeling of social processes* (pp. 97—125). London, UK: Sage.

Nowak, A., Szamrej, J. & Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review, 97*, 362—376.

Queller, S. & Smith, E. R. (2002). Subtyping versus bookkeeping in stereotype learning and change: Connectionist simulations and empirical findings. *Journal of Personality and Social Psychology, 82*, 300—313.

Read, S. J. & Miller, L. C. (1993) Rapist or "regular guy": Explanatory coherence in the construction of mental models of others. *Personality and Social Psychology Bulletin, 19*, 526-541.

Read, S. J. & Miller, L. C. (1998) *Connectionist models of Social Reasoning and Social Behavior*. New York: Erlbaum.

Read, S. J., & Montoya, J. A. (1999). An autoassociative model of causal reasoning and causal learning: Reply to Van Overwalle's critique of Read and Marcus-Newhall (1993). Journal of Personality and Social Psychology, 76, 728—742.

Riolo, R. L., Cohen, M. D., & Axelrod, R. (2000). Evolution of cooperation without reciprocity. *Nature, 414*, 441—443.

Rodriguez, M. A. & Steinbock, D. J. (2004). Societal-scale decision making using social networks. *North American Association for Computational Social and Organizational Science Conference proceedings*. Pittsburg: Pennsylvania at Carnegie Mellon University.

Ruscher, J. B. (1998). Prejudice and stereotyping in everyday communication. *Advances in Experimental Social Psychology, 30*, 241—307.

Ruscher, J. B. & Duval, L. L. (1998). Multiple communicators with unique target information transmit less stereotypical impressions. *Journal of Personality and Social Psychology, 74,* 392—344.

Schaller, M. & Conway III, L. G. (1999). Influence of impression –management goals on the emerging contents of group stereotypes: Support for a social-evolutionary process. *Personality and Social Psychology Bulletin, 25*, 7, 819—833.

Schober, M. F. & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*, 211—232.

Schuette, R. A. & Fazio, R. H. (1995). Attitude accessibility and motivation as determinants of biased processing: A test of the MODE model. *Personality and Social Psychology Bulletin, 21*, 704—710.

Schul, Y., Mayo, R. & Burstein, E. (2004). Encoding under trust and distrust: The spontaneous activation of incongruent cognitions. *Journal of Personality and Social Psychology, 86,*668—679.

Schultz-Hardt, S., Frey, D., Lüthgens, C. & Moscovici, S. (2000). Biased information search in group decision making. *Journal of Personality and Social Psychology, 78*, 655—669.

Shoda, Y., LeeTiernan, S., & Mischel, W. (2002) Personality as a Dynamical System: Emergence of Stability and Distinctiveness from Intra- and Interpersonal Interactions. *Personality and Social Psychology Review, 6*, 316—325.

Shultz, T. & Lepper, M. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological Review, 2*, 219-240.

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.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology, 70*, 893-912.

Spellman, B. A. & Holyoak, K. J. (1992). If Saddam is Hitler who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology, 62*, 913-933.

Spellman, B. A., Ullman, J. B., & Holyoak, K. J. (1993). A coherence model of cognitive consistency: Dynamics of attitude change during the Persian Gulf War. *Journal of Social Issues, 49*, 147-165.

Stasser, G. (1988). Computer simulation as a research tool: The DISCUSS model of group decision making. *Journal of Experimental Social Psychology, 24*, 393-422.

Stasser, G. (1999). The uncertain role of unshared information in collective choice. In L. L. Thompson, J. M. Levine & D. M. Messick (Eds.) *Shared Cognition in Organizations: The* *management of Knowledge* (pp 49—69). Mahwah, NJ: Erlbaum.

Stasser, G. & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*, 1467—1478.

Steels, L. (1999) *The Talking Heads Experiment*. Laboratorium: Antwerpen, Belgium

Steels, L. (2003). Evolving grounded communication for robots. *Trends in cognitive sciences, 7,* 308-312.

Steels, L. (2005). The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. *Connection science, 3-4,* 213-230.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and brain sciences, 28,* 469-529.

Sun, R. (2001). Cognitive science meets multi-agent systems: A prolegomenon. *Philosophical psychology, 14,* 5-28.

Sun, R., & Peterson, T. (1999). Multi-agent reinforcement learning: Weighting and partitioning. *Neural networks, 12,* 727-753.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435-467.

Thompson, M. S., Judd, C. M., Park, B. (2000) The consequences of communicating social stereotypes. *Journal of Experimental Social Psychology, 36*, 567-599.

Thorpe (1994). Localized versus distributed representations. In M. A. Arbib (Ed.) *Handbook of brain theory and neural networks* (pp. 949-952). Cambridge, MA: MIT Press.

Van Overwalle, F., Drenth, T. & Marsman, G. (1999). Spontaneous trait inferences: Are they linked to the actor or to the action? *Personality and Social Psychology Bulletin, 25,* 450-462.

Van Overwalle, F. & Labiouse, C. (2004). A recurrent connectionist model of person impression formation. *Personality and Social Psychology Review, 8*, 28—61.

Van Overwalle, F. & Siebler, F. (2005). A Connectionist Model of Attitude Formation and Change. Personality and Social Psychology Review, in press.

Van Overwalle, F. (1998) Causal Explanation as Constraint Satisfaction: A Critique and a Feedforward Connectionist Alternative. *Journal of Personality and Social Psychology, 74*, 312-328.

Van Overwalle, F., & Jordens, K. (2002). An adaptive connectionist model of cognitive dissonance. *Personality and Social Psychology Review, 6*, 204—231.

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.

Weiss G. (1999): Multiagent systems: a modern approach to distributed artificial intelligence. MIT Press: Cambridge, MA.

Wilkes-Gibbs, D. & Clark, H. H. (1992). Coordinating beliefs in conversation. Journal of Memory and Language, 31, 183—195.

Wittenbaum, G. M. & Bowman, J. M. (2004). A social validation explanation for mutual enhancement. *Journal of Experimental Social Psychology,* 40, 169—184.

Footnotes

[1] To control attenuation and boosting on an issue, we applied not only the *maximum*, but also tested the *average* of all receiving trust weights of all listeners on an issue. The maximum reflects the notion that if at least someone (or a *minority*) agrees with the talker, then he or she will not talk about the issue; the average reflects the same if the *majority* agrees with the agent. We did not test the *minimum*, or the notion that a talking agent would want to reach *unanimity* from the audience, as this seemed not very plausible. As the number of listeners in the present simulations was typically limited to only one or a few, the difference between maximum and average was negligible. However, we are aware that the desire to convince only a minority or a majority is very much under the strategic control of the speaker (see also *General Discussion*), so that it could be entered in the model as an additional parameter, especially when simulating a larger audience. We did not do so at present to keep the model conceptually coherent and as simple as possible.

[2] Each unit was replaced by a set of 5 units, and their activation was represented by a pattern drawn from a normal distribution with mean as indicated in the learning histories (i.e., Tables) and standard deviation 0.10, and this random pattern was redrawn after each 10th simulation run. In addition, to reflect the imperfect conditions of perception and inference, random noise drawn from a normal distribution with mean 1 and standard deviation 0.10 was added to these activations (hence, the variation in the starting weights was dropped). We kept the same parameters as for the localist representation, except that the standard learning rate was lower (as more units are involved) and the best fitting learning rate we found was either 0.03 or 0.06.

[3] The reader can also assess the robustness of the simulations by exploring other parameters. The files of the simulations are available upon request, and the program for running the simulations is available at www.vub.ac.be/PESP/VanOverwalle.html#FIT. Parameters can be changed and explored by choosing the menu *Simulation* | *Parameters*.

Table 1

Summary of the Main Simulation Steps for each trial in the TRUST Model

Cycle though steps A—G for all issues and their attributes (i.e., all units)	Equation in Text
A. set external activation for each issue object (or attributes) present (within agents)	*(1*)
B. spread that activation within all non-listening agent	*(2)* *(4b)* if agent ≠ *l*
C. attenuate and boost the internal activation of talking agents before transmitting it (see "i" in Tables 4 — 7)	*(7a)* *(7b)*
D. spread (i.e., transmit) activation from talking agents to listening agents (from "i" to "?" respectively in Tables 4 — 7)	*(6)*
E. spread that activation within all listening agents	*(2)* *(4b)* if agent = *l*
F. update trust weights (between agents)	*(8)*
G. update connection weights (within agents)	*(5)*

Note. "Within" refers to all units within an agent. During learning without communication, only Steps A, B, E and G are functional.

Table 2

Overview of the Simulations

Nr.	Topic	Empirical Evidence / Theoretical Prediction	Major Processing Principle
		Persuasion and Social Influence	
1	Number of Arguments	The more arguments heard, the more opinion shift	Information transmission leads to changes in listener's net [a]
i	Polarization	More opinion shift after group discussion	More information transmission by a majority
		Communication and Stereotyping	
2	Referencing	Less talking is needed to identify objects	Overactivation of talker's and listener's net
i i	Word use	Acquiring word terms, synonyms and ambiguous words	Information transmission on word meaning and competition between word meanings [a]
3	Stereotypes in Rumor Paradigm	More stereotype consistent information is transmitted further up a communication chain	Prior stereotypical knowledge of each talker and novel information combine to generate more stereotypical thoughts [a]
4	Perceived Sharedness	Less talking about issues that the listener knows and more talking about other issues	Attenuation vs. boosting of information transmission if receiving trust is high vs. low
5	Sharing Unique Information	Unique information is communicated only after some time in a free discussion	Same as Simulation 4

[a] The maxim of quantity (attenuation and boosting) did not play a critical role in these simulations.

Table 3

Principal parameters of the TRUST Model and features or individual and group processing they represent

Parameters	Human Features
Parameters of individuals nets	
Learning rate = .30	How fast new information is incorporated in prior knowledge
Starting weights = .00 ± .05	Initial weights for new connections
Parameters of communication among individual nets	
Trust learning rate = .40	How fast the trust in receiving information changes
Trust tolerance = .50	How much difference between incoming information and own beliefs is tolerated to be considered as trustworthy
Trust starting weights = .40 ± .05	Initial trust for new information

Note. ± .05 denotes a random number between -.05 and +.05 that was added to introduce some noise in the starting weights.

Table 4

Persuasive Arguments (Simulation 1)

	Talking Agent	
		Listening Agent
		Topic
	Feat1	Feat2
	Feat3	Feat4
	Topic	Feat1
	Feat2	Feat3
	Feat4	
	Prior Learning of Arguments	
#10		1
	1	1
	1	1
	Talking and Listening	
#1-3-5-7-9	1	i
	i	i
	i	?
	?	?
	?	?

Test of Beliefs

of Listener

		1
	?	?
	?	?

Note. Schematic version of learning experiences along Ebbesen & Bowers (1974, Experiment 3). Each row represents a trial (i.e., act) of learning or communicating where cell entries denote external activation and empty cell denote 0 activation; #=number of times the trial is repeated; i=internal activation generated by the talking agent after activating the discussion topic; ? = (little ear) external activation received from the talking agent; in the test phase ? = denotes the activation read off for measuring activation. Trial order was randomized in each phase and condition.

Table 5

Referencing (Simulation 2)

		Talking Agent		
			Listening Agent	
	Picture Martini	Glass Legs Each-Side Glass Legs Each-Side	Picture Martini	
		Prior Observation of Figure by "Director"		
#10			1	1
		1	.8	.4
		Talking and Listening		
#12			1	i
		i	i	i
		?	?	?
		?	?	
#8			?	?
		?	?	?
		1	i	i
		i	i	

Test of Talking

of “Director”		?	?
	?	?	
of “Matcher”			
	?	?	?
	?		

Note. Schematic version of learning experiences along Schober & Clark (1989, Experiment 1). Each row represents a trial of learning or communicating where cell entries denote external activation and empty cell denote 0 activation; #=number of times the trial is repeated; i=internal activation generated by the talking agent after activating the picture;? = (little ear) external activation received from the talking agent, in the test phase ? = activation of talking agents during the previous talking phase. Trial order was randomized in each phase and condition.

Table 6*: Rumor Paradigm (Simulation 3)*

Talking Agent

Listening Agent

Jamayans Smart Stupid

Honest

Liar

Jamayans Smart Stupid

Honest

Liar

Prior SC Information on Jamayans: Per Agent

#10 smart 1 1

#10 honest 1

1

Prior SI Information on Jamayans: Per Agent

#10 stupid

1

1

#10 liar

1

1

Mixed (SC + SI) Story to Agent 1

#5 smart 1 1

#5 liar 1

1

Talking and Listening by Agents 1→2, 2→3, 3→4, and 4→5

#5 intelligence 1 i

i

?

? ?

#5 honesty 1

i

i ?

? ?

Test of Talking by Each Agent

smart		
	?	
stupid		
		?
honest		
	?	
liar		
		?

Note. Schematic version of learning experiences along Lyons & Kashima (2003, Experiment 1). Each row represents a trial of learning or communicating where cell entries denote external activation and empty cell denote 0 activation; SC=Stereotype Consistent; SI=Stereotype Inconsistent; #=number of times the trial is repeated; i=internal activation generated by the talking agent after activating the Jamayans;? = (little ear) external activation received from the talking agent, in the test phase ? = activation of talking agents during the previous talking phase. The shared condition is always preceded by the SC Information for each agent, and the unshared condition is preceded by the SC or SI Information alternatingly for each agent, both followed by the Mixed Story, and Talking and Listening Phase. Trial order was randomized in each phase and condition.

Table 7

Shared vs. Unique Information (Simulation 5)

Talking Agent

Listening Agent

Patient Shared1 Shared2 Uni1 Uni2 Patient Shared1 Shared2 Uni1 Uni2

Learning the Medical Case from Video Tape

#5 1 1

1 1 0

#5

1 1 1

0 1

Talking and Listening

Shared 1 i

i

? ? ?

? ?

?

1 i i

Unique		1	
		i	i
	?		
	?	?	
		?	
		?	?
	1		
	i	i	
	Test of Talking		
			?
	?	?	?
		?	?
	?	?	

Note. Schematic version of learning experiences along Larson et al. (1996). Each row represents a trial of learning or communicating where cell entries denote external activation and empty cell denote 0 activation; Uni=Unique; #=number of times the trial is repeated; i=internal activation generated by the talking agent after activating the patient;? = (little ear) external activation received from the talking agent, in the test phase ? = activation of talking agents during the previous talking phase. Trial order was randomized in each phase and condition.

Figure Captions

Figure 1. A multi-agent network model of interpersonal communication. Each agent consists of an auto-associative recurrent network, and the communication between the agents is controlled by trust weights. The straight lines within each network represent intra-individual connection weights linking all units within an individual net, while the arrows between the networks represent inter-individual trust weights (only some of them are shown).

Figure 2. The functional role and adjustment of trust weights in communication between agents, illustrated when a talking agent expresses his or her opinion on the honesty (attribute) of the people of the Jamayans (issue object); if the talker's expressed activation is close vs. different from the listener's internal activation on the same attribute, this may lead to an increase vs. decrease of the trust weight involving that attribute (see text for more details).

Figure 3. Simulation 1: Attitude Shifts in function of the Number of Arguments heard. Human data are denoted by bars, simulated values by broken lines. The human data are from Figure 1 in "Proportion of risky to conservative arguments in a group discussion and choice shift" by E. B. Ebbesen & R. J. Bowers, 1974, *Journal of Personality and Social Psychology, 29*, p. 323. Copyright 1974 by the American Psychological Association.

Figure 4. Interlude Simulation i: Polarization in function of Progress in the Discussion and Amount of Communication. Decreasing polarization as the discussion unfolds is illustrated from left to right panel when communication between majority and minority groups is equal as within the groups [Left], is reduced to half that amount [Middle] or is totally cut off [Right] while keeping the overall communication alike.

Figure 5. Simulation 2: [Top] Referencing. In Krauss, R. M. & Fussell, S. R. (1991). Constructing shared communicative environments. In L. Resnick, J. Levine, S. Teasley (Eds.) *Perspectives on socially shared cognition*, p. 186. Copyright 1991 by the American Psychological Association. [Bottom] Words per Reference by the Director and Matcher. Human data are denoted by bars, simulated values by broken lines. The human data are from Figure 2 in "Understanding by addressees and overhearers" in M. F. Schober & H. H. Clark, 1989, *Cognitive Psychology, 21*, p. 217. Copyright 1989 by Academic Press.

Figure 6. Interlude Simulation ii: Lexical Acquisition for new, matched, synonymous and ambiguous words.

Figure 7. Simulation 3: Proportion of Stereotype-Consistent (SC) and Stereotype Inconsistent (SI) Story Elements in function of the Actual Sharedness. Human data are denoted by bars, simulated values by broken lines. The human data are from Figure 2 (averaged across central and peripheral story elements) in "How are stereotypes maintained through communication? The influence of stereotype sharedness" by A. Lyons & Y. Kashima, 2003, *Journal of Personality and Social Psychology, 85*, p. 995. Copyright 2003 by the American Psychological Association.

Figure 8. Simulation 4: Proportion of Stereotype-Consistent (SC) and Stereotype

Inconsistent (SI) Story Elements in function of Perceived Sharedness. Human data are denoted by bars, simulated values by broken lines. The human data are from Figure 1 in “How are stereotypes maintained through communication? The influence of stereotype sharedness” by A. Lyons & Y. Kashima, 2003, *Journal of Personality and Social Psychology, 85*, p. 995. Copyright 2003 by the American Psychological Association.

Figure 9. Simulation 5: Percent Shared Unique Information in function of Discussion Position. Human data are denoted by bars, simulated values by broken lines. The human data are from Figure 1 in "Diagnosing groups: Charting the flow of information in medical decision-making teams” J. R. Larson, Jr., C. Christensen, A. S. Abbott & T. M. Franz, 1996, *Journal of Personality and Social Psychology, 71,* p. 323. Copyright 1996 by the American Psychological Association.

.

Figure 1

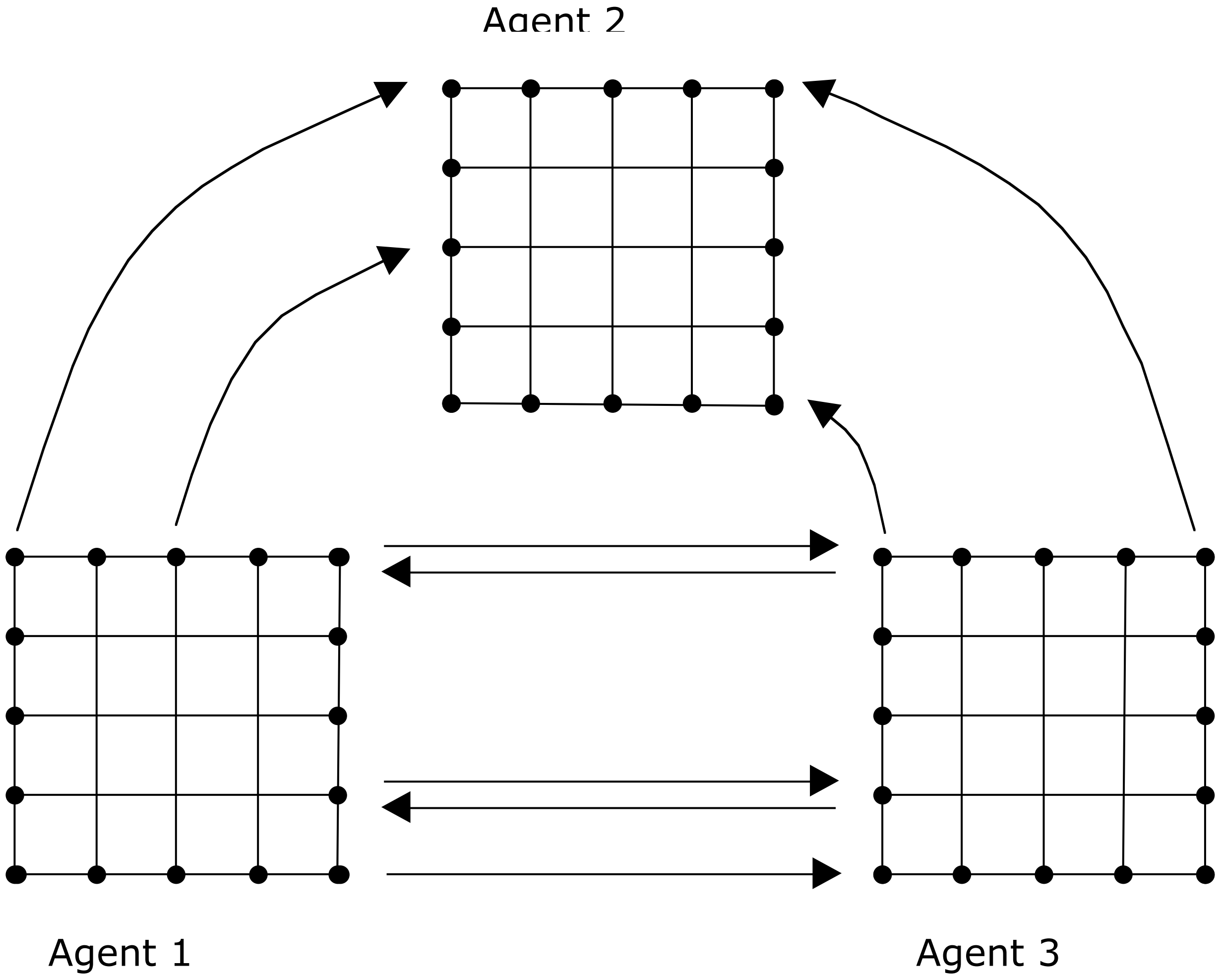
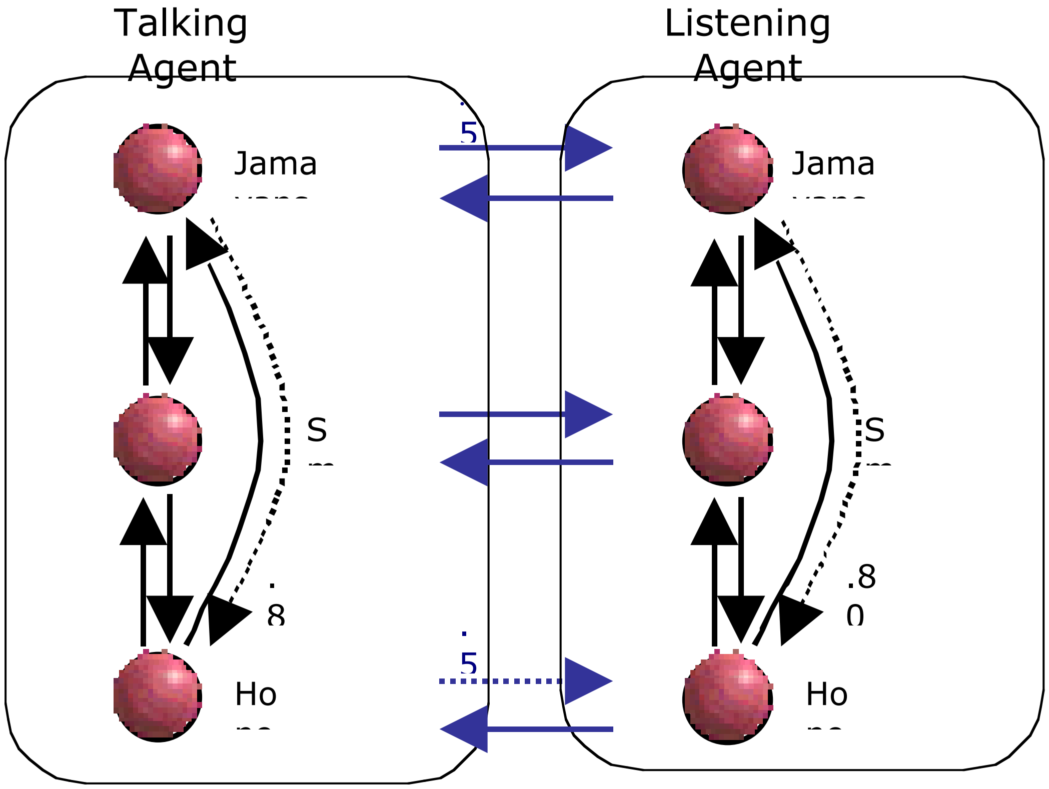


Figure 2

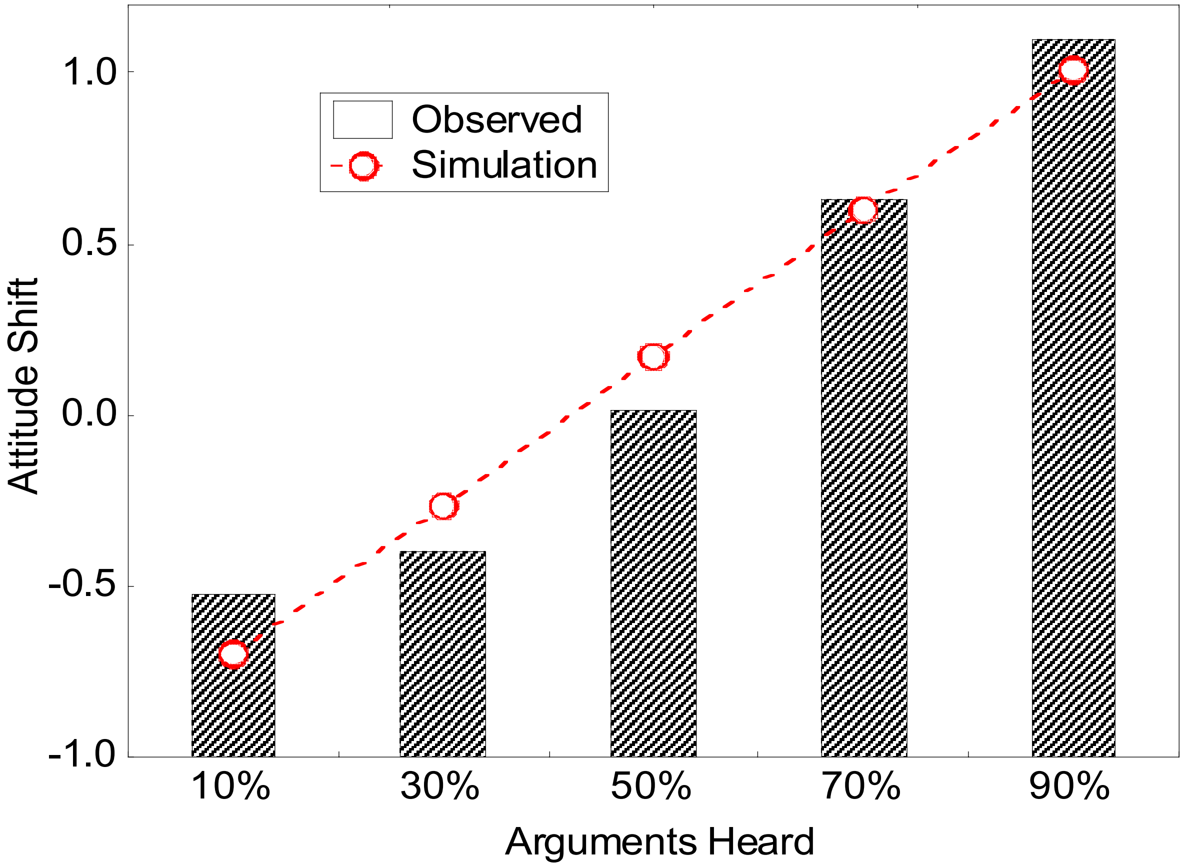


→ *Sending trust weight* $_{t\rightarrow l}$: Regulates how much the information sent by the talker is taken in

← *Receiving trust weight* $_{t\leftarrow l}$: Regulates how much the talker expresses the (novel vs. old)

⇢ *Trust weight* undergoing adjustment by comparing the external activation received from the talker (via talker's Jamayans→Honest internal weight and Honest→Honest trust weight) against the internal activation generated by the listener (via

Figure 3


Observed
Simulation
1.0
0.5
0.0
-0.5
-1.0
Attitude Shift
10%
30%
50%
70%
90%
Arguments Heard