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

Frank Van Overwalle, Francis Heylighen & Margeret Heath
Vrije Universiteit Brussel, Belgium

This research was supported by Grant XXX of the Vrije Universiteit Brussel to Frank Van Overwalle. We are grateful to XXX for their 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.

Running Head: TRUST: a Connectionist Model of Communication
[PUBTRUST]

7 June, 2005

Abstract

How is information transmitted in a group? A multi-agent connectionist model is proposed that combines features of standard recurrent models to simulate the process of information uptake, integration and memorization within individual agents, with novel aspects that simulate the communication of beliefs and opinions between agents. A crucial aspect in belief updating based on information from other agents is the trust in the information provided, implemented as the consistency with the receiving agents' existing beliefs. Trust leads to a selective propagation and thus filtering out of less reliable information, and implements Grice's (1975) maxims of quality and quantity in communication. By studying these communicative aspects within the framework of standard models of information processing, the unique contribution of communicative mechanisms beyond intra-personal factors 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 a deeper insight of its social origin. Sociologists and developmental psychologists have long noted that most of our knowledge of reality is the result of a social construction and communication rather than of individual observation. ~~Economists emphasize that effective knowledge management and learning is an organizational phenomenon that determines enterprise's success or failure (Senge, 1990).~~ Collective behaviors have a long evolutionary past, as biologists and computer scientists have built models that demonstrate how collectives of simple agents, such as ant colonies, bee hives, or flocks of birds, can process complex information more effectively than single agents facing the same tasks (Bonabeau et al., 1999).

Social psychologists too have studied cognition at the group level, using laboratory experiments to document various biases and shortcomings of collective intelligence. Research has revealed that we often fall prey to biases and simplistic stereotypes about other groups, and that many of these distorted representations are emergent properties of the cognitive dynamics in single or multiple minds. Some of these biased processes are *illusory correlation* or the creation of an unwarranted association between a group and undesirable characteristics, *accentuation* of differences between groups, *subtyping* of deviant members (for a review see Van Rooy, Van Overwalle, Vanhoomissen, Labiouse & French, 2004) and the communication of *stereotypes* (e.g., Lyons & Kashima, 1993). With respect to processes within a group, different types of social dynamics lead to a less than optimal performance. These include *conformity* and *polarization* which move a group as a whole towards more extreme opinions (Ebbesen & Bowers, 1974; Mackie & Cooper, 1984; Isenberg, 1986), *groupthink* that leads to unrealistic group decisions (Janis, 1972), 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, 2003) and the suboptimal use of relevant information channels in *social networks* (Leavitt, 1951; Mackenzie, 1976; Shaw, 1964).

These different approaches provide a new focus for the understanding of cognition that might be summarized as *collective intelligence* (Levy, 1997; Heylighen, 1999) or *distributed cognition* (Hutchins, 1995), that is, the cognitive processes and structures that emerge at the social level. To understand collective information processing, we must consider the distributed organization constituted by different individuals with different forms of knowledge and experience

and the social network that links them together and that supports their interindividual communication.

Multi-Agent Models

To develop a theory of distributed cognition, many researchers proposed multi-agent systems. In these systems, agents communicate, cooperate and interact in order to reach their individual or group objectives. The agents sometimes have different and conflicting knowledge and goals. However, the combination of their local interactions produces behavior at a higher collective level. According to Ferber (1989: cited in Bura et al., 1995, p. 89), an agent in such a distributed system should be

> a real or abstract entity that is able to act on itself and its environment; which has partial representation of its environment; which can, in a multi-agent universe, communicate with other agents; and whose behavior is the result of its observation, its knowledge and its interaction with the other agents (p. 249).

In spite of its promises, existing multi-agent approaches lacks a coherent framework that integrates the individual and the collective level of information processing. Previous systems such as cellular automata, social networks and many types of social neural networks lack some essential ingredients of a society of autonomous agents working and communicating together (for a fuller discussion, see section on *Alternative Models*). Perhaps the most crucial limitation of many models is that the individual agents lack their own psychological interpretation and representation of the environment. In fact, these models reduce each individual to a single unit or element possessing a rudimentary binary yes-no switch that denotes one's standing on an issue, rather than a human that exhibits complex and multifaceted thinking and reasoning. At most, the agents have a one-dimensional status that reflects their degree of belief, without any other mental capacities such as making links with other information, combining prior knowledge with current contextual information, evaluating the validity of incoming information and so on.

The present article aims to set the first steps towards an integrated theory combining the individual and collective level. Our approach is based on a limited number of assumptions: (a) groups of individual agents form a coordinated system that transmits information, (b) the resulting distributed cognitive system can be modeled as a connectionist network, and (c) information in the network is propagated selectively and gives rise to novel knowledge. 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 (see Figure 1). Each individual is represented by a recurrent auto-associative connectionist network, that is capable of representing internal beliefs as well as external information, and that can learn from its observations and memorize this. This type of recurrent model has been used 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).

Thus, we built on the strengths and capacities of a recurrent model as demonstrated by previous modeling research, to extent this approach to include communication between different individuals' networks. Communication involves the transmission of information on the same concepts from one agent's network to the other agent's network. In developing this extension towards multi-agent communication, we were strongly inspired by a number of communicative principles put forward by philosophers and psychologists (e.g., Grice, 1975; Krauss & Fussell, 1996). Two conversational principles of Grice (1975) that are very relevant for the present purpose are the *maxim of quality* ("Try to make your contribution one that is true", p. 46) and the *maxim of quantity* ("Make your contribution as informative as is required for the current purpose of the exchange", p. 45). (Other psychological aspects of communication incorporated in the model are discussed in the section on *Theories of Communication*.) In the following sections, we will see how the individual nets were created and how communication —including Grice's conversational maxims— was implemented in the model.

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 possessing adaptive learning mechanisms that develop 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 earning. In many algebraic models, 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, 1992). 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 outcome or 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.

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. Drawing on Marr's (1982) levels of information processing, earlier algebraic models are regarded the *computational* level of human reasoning which simply describe input-output relationships; connectionist models attempt to mimic psychological processes and therefore are considered the *algorithmic* level; and models that describe neural circuitry and processing that implement mental processes are regarded the *implementational* level. Although connectionist models are highly simplified versions of real neural functioning and only describe the algorithmic level of mental thinking, it is commonly assumed that they reveal a number of emergent processing properties that real human brains also exhibit. One of these emergent properties is that there is no clear separation between memory and processing as there is in traditional models. 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.

This article is organized as follows: First, we describe the proposed connectionist model in some detail, giving the precise architecture, the general learning algorithm and the specific details of how the model processes information within an individual as well as between individuals. We then give an overview of the simulations, and situate how they are related to existing theories of persuasive communication and conversation. This is followed by a series of simulations, using the same network architecture applied to a number of significantly different phenomena illustrating the different theoretical perspectives (see Table 3). We then discuss the implications and limitations of the proposed model and identify areas where further theoretical developments are needed. We end with a comparison of previous multi-agent models of collective behavior.

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). As noted earlier, this network has already been applied in social psychology to study, for instance, person and group impression formation (Smith & DeCoster, 1998; Van Rooy et al., 2003; Van Overwalle & Labiouse, 2004),

attitude formation and change (Van Overwalle & Siebler, 2005), and causal attribution (Read & Montoya, 1999). We apply this network for two reasons. First, we want to emphasize the theoretical similarities that underlie the simulations of these diverse social phenomena with the present findings of communication and the emergence of collective judgments and stereotypes. 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 (Shultz & Lepper, 1996; Siebler, 2002; 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 generic architecture of an auto-associative network is illustrated in Figure 2. Its most salient property is that all units are interconnected with all of 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 object (e.g., person, group, issue) as well as various attributes of the object that are associated with it. The connections linking the object with its attributes represent the individual's beliefs and knowledge about the object. For instance, an individual may belief that waitresses are talkative, and the strength of this belief is represented by the weight of the waitress→talkative connection. The units in the network can represent these concepts in basically two ways. In a localist representation, each unit represents a single symbolic concept like in earlier automatic spreading activation networks. In contrast, in a distributed representation, 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). Although a distributed representation is a more realistic neural code, for ease of presentation, we illustrate the basic workings of the recurrent model with a localist representation.

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 final 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 final activation is the linear sum of the external and internal activations after one updating cycle through the network. Previous simulations by Van Overwalle and colleagues (Van Overwalle & Labiouse, 2002; 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 employ a similar activation updating algorithm for our extension, 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

$$a_i = E * ext_i. \quad (1)$$

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

$$int_i = \sum_j (w_{j \to i} * a_j) \quad (2)$$

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

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

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

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

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

$$a_i = net_i = ext_i + int_i \quad (4b)$$

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 of an individual's weak waitress→talkative connection which creates a weak internal activation (e.g., the belief that waitresses are not talkative), the waitress→talkative connection weight is increased to reduce this discrepancy. Conversely, if the same external activation is overestimated because of an individual's waitress→talkative connection is too strong and creates 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 better approximate the external activation and to develop internal representations that accurately describe the environment. Thus, 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.

In mathematical terms, this error reducing capacity of the delta algorithm (McClelland & Rumelhart, 1988, p. 166) is formally expressed as:

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

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 increase to eventually reach an asymptotic value of +1 (see Figure 3). When this co-occurrence is not perfect, then the learning algorithm gradually converges to a connection weight that reflects the proportion of supporting and conflicting co-occurences.

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 connection weights linking the attributes between agents. A crucial aspect of these latter between-agents connections is that they reflect the degree of *trust*, or how much the information on a given object or attribute expressed by a talking agent is deemed 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 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 for each agent: Sending connections for a talking agent and receiving connections for a listening agent. These two trust connections implement to a great deal Grice's (1975) maxims of *quality* and *quantity* of communication.

Maxim of Quality: Sending Trust Weight

The maxim of quality suggests that in order to communicate efficiently, communicators generally try to transmit truthful information. In the model, this maxim of quality is implemented

on the side of the receiving agent. Communication is more efficient if the information is believed to be trustworthy. This is implemented in the trust connection from a talking agent expressing his or her ideas to the receiving agent. When trust is maximal (+1), the information expressed by the talking agent is unattenuated by the listening agent. To the degree that trust is lower, information processing by the listener is attenuated in proportion to the trust weight. When trust is minimal (0), no information is processed by the listening agent. This mechanism is schematically represented in Figure 4 (top arrow).

Thus, the listener *j* sums all information received from other talking agents *i* in proportion to the trust weights (and then processes this information internally according to the standard recurrent approach described above). Or, in mathematical terms:

$$ext_j = \sum_i (t_{i \blacklozenge j} * a_i) \qquad (6)$$

where ext_j represents the external activation received by the listening agent *j* from the talking agent *i* or the degree to which the external information is ready for processing by the listening agent *j*; $t_{i \rightarrow j}$ is the trust weight from the talking agent *i* to the listening agent *j*; and a_i denotes the activation expressed by talking agent *i*. By comparing with Equation 1, it can be seen that this mechanism of trust spreading between agents is a straightforward copy of activation spreading of connectionist models within a single agent. This suggests that, except for the transmission of information (by speech or other means), the acceptance of the information by the listener on the basis of feelings of trust is most probably a relatively automatic process. Recent neurological research (King-Casas et al., 2005) seems to support this view of automaticity of trust.

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. 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 trust weights from the listening agent to the talking agent (see Figure 4, bottom arrow). These weights indicate the degree of trust by the talking agent in the listening agent, and are determined by earlier communications in which the listening agent expressed judgments on an issue that were either congruent or incongruent with the talking agent's ideas. To the extent that these trust weights are high (above trust starting weights,

$t_\circ$), knowledge and agreement on an issue is assumed and the talking agent will restrain him- or herself from expressing these ideas further (attenuation). In contrast, when these weights are low (below trust starting weights, $t_\circ$), the talking agent tends to express and defend his or her ideas on this issue more strongly (boosting). In mathematical terms:

$$\begin{aligned} &\text{if } \max(t_{i\leftarrow j}) < t_\circ \\ &\text{then } a_i = a_i * [\, 1 + max(t_{i\leftarrow j}) \,] \\ &\text{else } \; a_i = a_i * [\, 1 - max(t_{i\leftarrow j}) \,] \end{aligned} \tag{7a}$$

where a_i is the activation expressed by the talking agent *i*, and $max(t_{i\leftarrow j})$ represents the maximum trust weight from all listening agents *j* to a talking agent *i*. In contrast, for all other issues *k* agent *i* is talking about, the reverse change of activation occurs:

$$\begin{aligned} &\text{if } \; max(t_{i\leftarrow j}) < t_\circ \\ &\text{then } a_k = a_k * [\, 1 + max(t_{i\leftarrow j}) \,] \\ &\text{else } \; a_k = a_k * [\, 1 - max(t_{i\leftarrow j}) \,] \end{aligned} \tag{7b}$$

Because this mechanism of attenuation and boosting of activation (or expression of information by a talking agent *i*) is a rather novel concept in theory construction, it is unclear whether this is either a largely automatic process or a more controlled strategy used by the speaker.

Adjustment of Trust Weights

Given that perceived trust plays a crucial role in the transmission of information, it is important to describe how trust is developed and changed in the model. Like the standard delta learning algorithm which is used to adjust memory traces within individual agents, the degree of trust depends on the error between external beliefs expressed by a talking agent and a listening agent's own internal beliefs. If the error is below some *trust tolerance*, the trust weight between the concepts held by the two agents is increased towards 1; otherwise, the trust weight is decreased towards 0. In mathematical terms, the trust weight change of the listening agent j in the beliefs expressed by agent i, or $\Delta t_{i\rightarrow j}$, is implemented as follows:

$$\begin{aligned} &\text{if } | \, ext_j - int_j \, | < trust\ tolerance \\ &\text{then } \Delta t_{i\rightarrow j} = \gamma * (1 - t_{i\rightarrow j}) * | \, a_i \, | \\ &\text{else } \; \Delta t_{i\rightarrow j} = \gamma * (0 - t_{i\rightarrow j}) * | \, a_i \, |, \end{aligned} \tag{8}$$

where ext_j represents the external activation received (from the talking agent *i*) by the listening agent *j* and int_j the internal activation generated independently by the listening agent *j*; and γ is the

rate by which trust is adjusted and $| a_i |$ the absolute value of the activation of the talking agent *i*. In contrast to Equation 5 of the delta algorithm, in Equation 9, 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).

Because the trust change mechanism is a straightforward extension of the basic delta learning algorithm in that it is also error-driven and attempts to reduce the error between the listening agent's internal representation and external information, we assume that this mechanism is largely automatic.

Summing up, the larger $t_{i \rightarrow j}$ becomes, the more the listening agent *j* will trust the talking agent *i* on the issues communicated, and the more influential the talking agent will become (maxim of quality). In turn, this will restrain the just-listening agent *j* in expressing his or her ideas on this issue (maxim of quantity). Note that when a listening agent's own beliefs are changed as a result of the feedback from some agents, this will have an effect on the listener's own internal activation (int_j) and so on his or her perceived trustworthiness of all other agents. A summary of the steps in the simulation of a single learning 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 3 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.

We first describe the successive learning phases in the simulations and how cognitions were coded in the network, how beliefs or conversational content were measured, and we end with the general parameters of the model.

Learning Phases

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 objects with some characteristics. 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 one or more agents communicated. As we will see shortly, talking units were construed by activating the target object and letting the activation spread to the associated characteristics in the agent's net. Thus, both the topic of the conversation as well as the internally generated thoughts on its associated characteristics by the speaker was transmitted to the listener. Given that the topic was typically imposed externally (by the experimenter), this was designed as external activation, while the participant's own thoughts were designated as internal activation (denoted by "i" in the tables) generated after activating the topic. Communication was then simulated by spreading this external and internal activation via the trust weights to the corresponding listening agents' units that represented the same concepts (denoted by "?" — a little ear— in the tables). The exact order of the computations is detailed in Table 2. 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). Nevertheless, the major results hold across a wide range of stimulus distributions that we tested.

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. In one set of experiments, the measures involved judgments or opinions by the participants, or their memory of the arguments provided. We assume that the task instructions activate the target object in participants' mind, and that they then react to the resulting activation of the associated characteristics of interest. Thus, for instance, if arguments

are given on a certain attitude topic and participants are later on requested to provide their opinion, in the network, this request spontaneously activates the topic and after automatic activation spreading in the agent's net, the resulting mean activation of the arguments provides an indication of the agent's global belief. To simulate such *Test of Beliefs* in the agents' networks, we turned the object unit (i.e., object) on and recorded the resulting activation in the relevant feature units (i.e., arguments; see also Equations 2 & 3).

In another set of experiments, the content of the communicated messages was recorded and served as dependent measure. To simulate such *Test of Talking*, we recorded the mean activation of each unit of interest during the preceding Talking and Listening Phase. Hence, this measure reflects the average amount of activation of a given object or its features, that is, how strongly these were expressed by the talking agents.

These 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

For all simulations, we used the linear auto-associative recurrent network described above. In spite of the fact that the major experiments to be simulated used very different stimulus materials, measures and procedures, all the model parameters are set to the same values, unless noted otherwise. Specifically, the parameters of the individual nets were the same as in earlier simulations by Van Overwalle and colleagues (E = *Decay* = number of internal Cycles = 1, and a linear summation of internal and external activation; see also Van Overwalle & Labouise, 2004; Van Overwalle & Siebler, 2004). 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 2. In order to ensure that there was sufficient variation in the data to conduct meaningful statistical analyses, all connection weights were initialized at the specified values plus additional random noise between -.05 and +0.05.

Theories of Communication and a Preview of the Simulations

Because earlier multi-agent models did not take into account how information is represented and processed by an individual agent, virtually no attention was given to the conversational mechanisms by which information was actually transmitted, and relevant psychological research was largely ignored. However, there are a number of important theoretical perspectives and research findings in this domain. In their review of interpersonal communication theories, Krauss and Fussell (1996) summarized these in four distinct psychological approaches. Which aspects of these perspectives on communication are incorporated in the multi-agent system that we propose here?

The first approach is labeled the *encoder-decoder* perspective. It assumes that communication is a simple transmission of signals, and that there is basically a single meaning for each signal. The speaker's mental representation is transformed into a verbal representation (e.g., speech) by an encoder, and the listener is able to understand the message by a decoder that recreates a mental representation that corresponds to the representation of the speaker. Differences between the mental representations of the speaker and listener may exist because of, for example, different understandings of the symbols used in the messages, incorrect decoding of the mental representation implied by the speaker and so on. To represent human communication, any model must be minimally capable to account for a transmission of representations. The proposed network does so, although it leaves unspecified the encoding and decoding process by which the individual representations are transformed in a verbal format. Simulations 1 and 2 on persuasive communication are included in this article to demonstrate that the simple transmission of mental representations between individuals is possible in the model, although this transmission is governed by additional factors discussed next.

A second approach is the *intentionalist* perspective. In addition to message transmission, it focuses on the process of inference by which the underlying communicative intent of the speaker is derived. To do so, Grice (1975) proposed that people see a conversation in a cooperative light. Even when the underlying goal is to criticize or compete against the other person, in order to get the message across, people must collaborate to shape a meaningful message. Grice (1975) derived from this cooperative principle two important conversational maxims: the *maxim of quality* ("Try to make your contribution one that is true", p. 46) and the *maxim of quantity* ("Make your contribution

as informative as is required for the current purpose of the exchange", p. 45). These two maxims have gained increasing attention in social psychology (e.g., Lyons & Kashima, 2003) and are also incorporated in the present model. All the simulations make use of the maxim of quality —the cornerstone of our model— while Simulation 5 (on transmission of stereotypes) and 6 (on the use of unique information) specifically target the application of the maxim of quantity. However, the proposed model does not address other aspects of the intentionalist perspective, such as the idea that messages may imply several acts besides transmission of meaning, such as the emission of demands, promises or other requests for a (behavioral) response from the listener (see Searle, 1979).

To establish a relevant context from which the meaning of utterances can be interpreted, it is necessary to create a common ground or mutual knowledge among the conversants. This issue is addressed by the *perspective-taking* view. It presupposes that speakers and listeners have each a somewhat divergent vantage point from which they interpret the communication. To bridge these differences, conversants must tailor their speech to the listeners and common background knowledge must be created. This process is termed *grounding* (Schober & Clark, 1989). However, theorists differ on many aspects. How much mutual knowledge is required —just about the current conversation topic, or is knowledge about beliefs and opinions of the listeners also necessary? Is a full fletched model of the other individual necessary, or is momentarily feedback sufficient to understand the current message? Recent findings seem to suggest that to create a message it is not necessary to rely on a model of the listener's knowledge constructed from prior assumption, and that momentarily feedback might be sufficient as it enables immediate correction of misunderstandings. This issue is addressed in research using the so-called *referential* paradigm, in which one person describes one item in an array in such a way that the other person can identify it (e.g., Kingsbury, 1968; Krauss & Weinheimer, 1964, 1966; Schober & Clark, 1989; Steels, 1999; Wilkes-Gibbs & Clark, 1992). Simulation 3 replicates this paradigm and the simulation results point out that no explicit model of the other agent is needed, save for the memory of the agent's trustworthiness on some specific topics (which is part of the proposed model). However, the model presented does not address other aspects of the perspective-taking approach such as the finding that speakers tend to bias their description of a target's traits in the direction of the listener's attitude (McCann, Higgins & Fondacaro, 1991; Schaller & Conway III, 1999).

According to the final, *dialogic* perspective, a communication exchange is more than the combined output of two autonomous agents, but rather a joint accomplishment that is socially situated so that the meaning of the message can be understood only in a particular context. For instance, in the referential paradigm, it has been found that only participants of the conversation profit from the common ground created by them, that overhearers who enter in the middle of a conversation (but heard the whole conversation) lack the shared knowledge in that they did not co-create a common background and thus takes much less advantage from it (e.g., they might have misunderstood some point in the conversation but could not give feedback to correct). This aspect is also addressed in Simulation 3.

The simulations we describe next address to main themes. The first section on *Persuasion and Social Influence* involves the exchange and change of beliefs through communication, and requires a minimal set of assumptions for the transmission of agents' representations states. The second section on *Communication* zooms in on various conversational principles such as common ground and referencing, with important consequences on stereotyping and sharing of information.

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 possible 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 simulations demonstrate that the number of arguments provided to listeners are crucially important in changing their beliefs and that the sources of these arguments determine the trust in this information and ultimately regulate its impact. Additionally, we illustrate that the TRUST multi-agent model predicts group polarization.

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. To demonstrate this effect of size, participants listened to a tape-recording of a group discussion. The recording contained a range from little (10 %) to many (90 %) arguments in favor or a more risky choice. As can be seen in Figure 5, 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 topic of discussion, and four units to reflect the features of the arguments. As can be seen, the topic and feature 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.
- Next, the discussion group talks about the issue. As can be seen, talking is implemented by activating the topic of discussion in the talking agent and then allowing the agent's internal activation (i.e., own beliefs) generate external activation. This external activation thus reflects the "talking" about one's own opinions. This activation is then spread to the listening agent in proportion to the trust weights where it is received (as indicated by a "?" [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 that 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 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 5, 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 result do not change when attenuation and boosting (maxim of quantity) is turned off in the simulation.

Simulation 2: Trust in Source

Having demonstrated how information can be passed from one agent to another, it is important also to illustrate the crucial role of trust. There are several determinants of trust, 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 much more than members of another group. Based on this assumption, the TRUST model predicts that information from ingroup members has a stronger effect on listeners than in formation from an outgroup member.

Key Experiment. The effect of group membership on the impact of persuasive messages was investigated by Mackie and Copper (1984, Experiment 1), using the same paradigm as Ebbesen ad Bower (1974). Participants listened to a tape-recording that contained arguments in favor of retention or abolition of standardized tests for university entry. As can be seen in Figure 6, students' attitudes towards the tests was dramatically altered after hearing the tape from an allegedly ingroup, but were much less affected when the tape was alleged to be from a outgroup. Mackie (1986) found the same effect even without outgroup categorization, as it was sufficient to describe the "outgroup" tapes to be from a collection of unrelated individuals to eliminate its persuasive impact.

Simulation. Table 5 represents a simplified learning history of this experiment. The architecture is identical to the previous simulation, and the simulation history is very similar. There are a few differences. First, the talking agents learn features of the arguments that are not only in favor of an opinion (i.e., retention of the entry tests), but in some conditions also against it. Second, they all express their arguments 5 times instead of a varying number of times. The most important novelty of this simulation is the manipulation of trust. In the ingroup condition, all talking→listener trust weights are set to +1 before the simulation to reflect that the listeners trust the talking agents completely. Conversely, in the outgroup condition, all talking→listener trust weight are set to 0 to reflect complete lack of trust.

Results. The "statements" listed in Table 5 were processed by the network for 50 "participants" with different random orders. In Figure 6, the simulated values (broken lines) are compared with the attitude shifts (striped bars) observed by Mackie and Copper (1984, Experiment 1). As can be seen, the simulated and observed data match quite well and the correlation between them is significant, $r = .95$, $p < .05$ (one-sided). An ANOVA on the simulated attitude further reveals a significant interaction between Arguments (pro vs. anti) and Group (ingroup vs. outgroup), $F(1,196) = 16.80$, $p < .001$. Separate t-tests show that the difference between pro and anti arguments is significant for the ingroup, $t(98) = 5.08$, $p < .001$, while it is not for the outgroup, $t(98) = 1.91$, $p = .06$ (and in the wrong direction). Note again that these results do not change when attenuation and boosting was turned off in the simulation.

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 each agent during several discussion rounds to exchange his or her ideas two times with each of the other agents. Thus, on each discussion round, one agent talks at the time while another agent listens, and this goes on until all agents talk two times to everybody. 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 Figure 7, polarization was already obtained after one discussion round as the groups average position moved in the direction of the majority position, and 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 a larger default value of 0.50, all agents 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 is 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 Simulations 2 and 3.

Communication

Communication is a primary means by which people attempt to influence and convert each other (Kraus & Fussell, 1996, 1991; Ruscher, 1998). In studying how communication between people is accomplished, researchers have developed several empirical paradigms. Some paradigms focus on language itself or non-verbal and paralinguistic aspects of conversation. However, 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. One of such paradigm is *referencing*, or the use of language to describe and identify some state of the external environment such as concrete of abstract objects, or even feelings and appreciations. This gives an opportunity to study the process by which people coordinate and establish a joint perspective during the conversation (Schober & Clark, 1989). Another paradigm takes its inspiration from rumors, and how they are sequentially spread in a community (Lyons & Kashima, 2003), while still another paradigm explores free discussions and how information that is unique to some members is shared

in the whole group (Tasser, 1999).

Simulation 3: Referencing

Key Experiment. Several studies explored how people use language to identify abstract or concrete objects outside them (e.g., Kingsbury, 1968; Krauss & Weinheimer, 1964, 1966; Schober & Clark, 1989; Steels, 1999; Wilkes-Gibbs & Clark, 1992). In research on referencing, typically, one person is designated as the "director" and is given the task to describe a number of 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. As we have seen earlier, this collaborative process is also termed *grounding*. The aim of the research is to asses how this perspective-taking and coordination influences the participant's messages and the adequacy of these messages. Figure 8 (top panel) depicts the typical outcome of such studies (Kraus & Fussell, 1991). On their first reference to one of these pictures, most directors use a long description, consisting of pictorial descriptions. 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, this 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 simple 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 & Clark, 1989).

Schober and Clark (1989) conducted an interesting experiment because they recorded not only the verbal descriptions of the director, but also the reactions of the matcher (Experiment 1). In addition, they measured the accuracy on the identification task of the matcher and of another participant that only overheard the conversation right from the beginning (early overhearer) or later while the conversation had elapsed for some time (later overhearer). The results of the analysis of the messages (see Figure 8, bottom panel) show 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. Figure 9 plots the percentage correct answers. It can be seen that although both matchers and (early) overhearers have access to the same messages, the matchers take most advantage of this information because their questions are tailored to their own information needs and doubts. Late overhearers are even more handicapped, because they

missed most of the initial descriptions by the director.

Simulation. Table 6 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 8 (top panel). Moreover, to reproduce a natural conversion, we allowed both agents to talk and listen one after the other in a randomized sequence, with the sole limitation that the director talked more often than the matcher. 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.

The accuracy of the agents (Figure 9) was tested like in the previous simulations, by priming the object and letting the activation spread to the features. Our assumption is that the stronger these activations, the better the participants can activate the correct image, and so provide correct answers. To test the content of the messages (Figure 8), we simply measured the average activation during the Talking and Listening Phase just prior to the Test Phase.

Results. The "statements" listed in Table 6 were processed by the network for 50 "participants" with different random orders. In Figure 8 (bottom panel), the simulated values (broken lines) are compared with the 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, $F(5,294) = 56.26$, $p < .001$. Figure 9 shows the accuracy observed by Schober and Clark (1989) compared with the simulation results for the accuracy data. There is again a close fit and significant correlation between simulated and observed data, $r = .91$, $p < .01$. Note, however, that these simulations are much less robust (especially after the introduction of random starting weights), which may indicate a weakness in the simulation or perhaps that the empirical data can differ strongly between participants and studies. To our knowledge, however, there are no other studies with accuracy data, so this latter assumption cannot be verified.

Extensions and Discussion. The same simulation approach was applied with success on related work on referencing (Kingsbury, 1968; Krauss & Weinheimer, 1966, 1968) with the same parameters, except for a lower initial trust weight in Kingsbury (1968).

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. 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, 1989; Van Overwalle & Labiouse, 1993; Van Rooy et al. 1994). 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 drastically from the observed data.

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 relies on a

basic principle of cognitive economy that our brain uses in many instances.

Interlude Simulation: Talking Heads

In his "Talking Heads" experiment, Steels (1999) 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 gained meaning and developed languages. Although the present TRUST model obviously has not the full capacities of Steels' robots, it is able to replicate the naming game that the robots also played. Four different sort 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.

To simulate these four types of meaning creation, a talking an 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 creates 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 10 displays how the meaning of the words are created and changed under these four circumstances. Note that we set 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 his high strength it has from the beginning. For synonyms, the simulation reveals a competition between words. To denote a circle, the listening agent gradually looses 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") although the listening agent keeps his initial preference for Square→"xu" as if it is the more prototypical member of the category.

Simulation 4: 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 aggravated 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 disconfirmation 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, 19??). 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 on stereotypes is more efficient when the information is 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. In a serial chain paradigm, 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 of 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 11 (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 of 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 7, for the actual shared condition, we provided 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). Next, 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 7 for 50 "participants" in the network with different random orders. As can be seen in Figure 11, 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 5: 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 information can be attenuated, while the expression of other information can 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 12, 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) 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 largely agree with the speaker.

Results. As can be seen in Figure 12, 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 the implementation of the maxim of quantity. If only boosting on other issues was 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. Without it, the simulation revealed an almost identical pattern of SC and SI information as for the ignorance condition. This strongly suggests that the maxim of quantity helps to neutralize the stereotype confirmation bias in communication.

Simulation 6: Group Problem-Solving and Sharing Information

Decisions are often made in a group rather than left to individuals, because it is often believed that a group as a whole has more intellectual resources to help solve the problem. Some members may have crucial information that may lead to a solution and that others do not have. By pooling all such unique information available, the group as a whole can 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 does, it is often brought in the discussion much later. This, of course, reduces the efficiency and speed of group problem solving (Larson, Christensen, Abbott & Franz, 1996; Larson, Foster-Fishman & Franz, 1998).

However, 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 (1992, 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 its 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, a 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 videotapes were created for each patient and they differed from another with respect to the specific items of information they contained. Each tape contained some information that was also present in the other tapes (shared condition) while some was present in one tape alone (unique condition). Roughly half of the information was shared. Each videotape was seen by different team members. After having seen the videotaped examination, beginning (students) and established medical experts discussed the case as a team and produced a differential diagnosis. This discussion typically lasted less than 20-25 minutes. Figure 13 shows the percentage of shared as opposed to unique information as the discussion unfolded. As can be seen, there is a negative linear trend in that initially a lot of shared information is discussed, while at the end more unique information is mentioned.

Simulation. We simulated a discussion by 3 agents (although only 2 are shown in Table 8). Each agent had 7 units (although only 5 are shown) consisting of the topic of the information exchange (i.e., the patient), and 3 shared items and 3 unique items. To simulate the viewing of the videotape, we ran 5 trials for each agent, in which all the shared items were learned and a single unique item. Next, we let all agents freely talk about all the shared and unique items. However, 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. After each discussion round in which each agent expressed a single shared and a single unique item, we measured how much each agent had communicated the shared and unique items.

Results. We ran the "statements" from Table 8 for 50 "participants" with different random orders. As can be seen in Figure 13, the simulation closely matched the observed data ($r = .86$, $p < .001$). A one-way ANOVA on the percentage shared information reveals the predicted main effect of the discussion position, $F(34, 2665) = 409.80$, $p < .001$. 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 to 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: flowchart-like models, 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.

Flowchart Models

Like a computer flowchart or a decision tree, flowchart models describe the different decision steps through which people maneuver in order to make rational decisions. These decision steps are often directly derived from empirical results in which important factors determining human behavior and decision making were uncovered. For instance, in their model of jury decision, Penrod and Hastie (1980) use parameters such as jury size, requisite verdict-rendering

number of votes, resistance to persuasion, time limit, value of postdecision increase in juror's persuasion resistance, and so on, as parameters in the model, and describe a flowchart in which each of these parameters determine the next step in the decision process. No doubt, these models make the empirical determinants of human behavior more explicit, and allow studying the interaction between these determinants in more detail. However, they are strongly limited by their lack of an underlying psychological theory supporting the decision structure and parameters of the model, by the assumption that opinions and beliefs are only developed at an explicit level of reasoning, and by the fact that they describe rather than explicate the decision process.

Cellular Automata and Social Networks

While flow chart models attempt to reproduce people's thought processes —albeit in a rudimentary form— the set of models we discuss next are less concerned with process *within* an individual, but rather *between* individuals. In cellular automata and social networks, the units of the model represent single individuals and the connections the relationships between individuals.

Cellular Automata. Cellular automata consist of a number of units (automata) arranged in a regular spatial pattern such as a checkerboard. Each automaton can be in a limited 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 persuasiveness, propensity for cooperation, decision strategies and so on (Nowak et al., 1998). Cellular automata allow to capture in formal terms the regularities and patterns of social norms or opinions in a group. For instance, Barr (2004) demonstrated that a whole population can converge to a single symbolic system (such as language) when individual agents update their behavior on the basis of local symbolic information between neighboring agents only, and if that would fail, spatially organized dialects emerge. Similarly, Nowak, Szamrej and Latané (1990) demonstrated how deviant opinions of minority groups survive in a whole population, and Couzin, Krause, Franks and Levin (2005) illustrated that only a small proportion of informed individuals is needed to guide a group of naïve individuals towards a goal. 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 each individual interacts only with his or her neighbors. 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). However, a limitation of social networks is that the links are often specified only in positive or negative 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 directly), 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 model 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 is represented by a single recurrent net (with hidden layers), and all the nets can exchange information with each other. The model illustrates how a common lexicon is created by sharing information between individuals. However, as we understand it, the output of one individual's network serves as direct input for another individual's network as if two brains are 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.

Implications and Limitations: What has been Accomplished?

In the preceding sections, we reviewed various simulations of existing empirical data which constitute "post"dictions of our TRUST model. Looking back at the theories of communication that inspired this empirical research (see section on *Theories of Communication)*, it is immediately evident that our model was capable to integrated and replicate many basic principles that these models put forward. However, it is also evident that we have charted only the first steps and incorporate these principles only in a modest form. Several aspects have been deliberately left out in order to focus on the most important social aspects of communication in a small group of individuals.

Linguistic and Other Media. It is clear that our model encompasses the transmission of information, although it does so mainly by leaving unspecified by what means information is simple encoded and decoded in a linguistic format. In most social psychology experiments, information transmission is simply accomplished by verbal means: speech or writing. How exactly the outputs of the semantic units in our connectionist system are transformed to speech or writing by the talking agent, and how this is again transformed to the input for the connectionist system of the listening agent is left out of our model. However, all our simulations demonstrate that this approach is sufficient to characterize the basic underlying principles of important communication phenomena. Thus, at least at the moment, the omission of language as a tool of communication seems a sensible simplification. Moreover, social communication may be driven by other media than verbal language, such as non-verbal behavior, emblems or sign language. Deaf sign language is a nice example demonstrating that some types of communication depends less on semantic symbols (which are not related to the concepts they denote), and more on visual signs which are more strongly related to the concepts they depict.

Sure, for a more complete investigation of intelligent social interaction and collaboration, the use of human media, be it verbal or non-verbal, is essential (Kraut & Higgins, 1984; Krauss & Fussell, 1991). Moreover, the role of external media and artifacts such as modern information technologies that support individual thought and interindividual communication is also of interest. Language is also an interesting avenue for further modeling. Some theorists already 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).

Cooperative Maxims and Perspective Taking. Perhaps our model is most powerful at the level of the representation and change of individuals' beliefs, as it explains how conversational maxims support the transmission of these representations and takes into account the individuals' perspective. In fact, the comparison with the agent's own perspective is the most critical determinant in our model to determine trust, which is the key in transmitting information. The role of Grice's maxim of quality (or trust weights) was made evident in all simulations, and the role of the maxim of quantity (or attenuation and boosting) was especially made clear when the perceived trust in the other agent was crucial (e.g., Simulations 5 & 6). One of the pressing questions in this line of investigation is to what extent a full blown internal model is needed about the listener in order for the speaker to create an efficient conversation. Our simulations make clear that such an explicit model is not needed, and that only one's trust in specific topics from each agent needs to be (implicitly) remembered. Although a full blown model is perhaps not needed, this does not imply that people do not develop or use them. Given that research has documented that inferring traits about others is almost inevitable (e.g., Van Overwalle, Drenth & Marsman, 1999), it may well be that it is developed used more often than required.

Social Strategies. Although the present model performs rather well with respect to perspective taking by individual agents, it largely ignores the social context and the social role of the communicators. This is most obvious when other motives play a role in the conversation beside Grice's maxims (McCann & Higgins, 1992). Thus, a speaker may attempt to develop or maintain a satisfactory interpersonal relationship or he or she may attempt to conform to social rules and norms in order to promote the self and to avoid public embarrassment. Such motives may have greater priority, even on an evolutionary scale, than transmitting truthful messages as implied by the maxim of quantity (Schaller & Conway III, 1999). Research has revealed that individuals strategically alter the content of their communications in response to impression management goals. For instance, communication tends to be biased in the direction of the audience when speakers believe that they can make a better impression by doing so or when the target person of a conversation is liked by the audience (McCann, Higgins & Fondacaro, 1991; Schaller & Conway

III, 1999). Other strategic concerns may involve to what extent the information is useful for the audience, to what extent the audience requires a simple versus complex understanding, a more formal versus informal description, and so on. Although such strategic motives are an interesting question on their own, it seems to use that they are often made deliberative or at least have to be learned during socialization. This indicates that they are often outside the implicit level of information transmission that the trust model attempts to simulate, at least in its current form. One can think of extending the model by modules that incorporate the contextual or task related goals, and as such give higher or lower priority (i.e., activation) to goal-relevant information.

The Societal Level. In comparison with earlier multi-agent models, it is obvious that the TRUST model does more justice to the complexity of the human mind, in that individuals are not reduced to single elements and have their own perspectives and opinions. However, by zooming in on the individual level, the model may have lost power on the larger societal level, which is the domain where earlier multi-agent models excel. It is unclear whether the TRUST model or any similar collection of individual agents will scale up so that it can explain phenomena at a larger scale of societies rather than agents.

Novel Hypotheses: What to do next?

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 such novel predictions. A first hypothesis tests whether the TRUST model is capable of developing limited or extensive internal models of other agents. The next hypotheses involve a number of empirical avenues for testing some core assumptions of our theoretical approach. As noted earlier, we consider the trustworthiness of information as theoretically most crucial in a communicative context, as well as the criterion of novelty. Although there is increasing awareness that trust is an important "core motive" in human interaction (Fiske, 2005) and while its neurological underpinning are gradually unveiled (King-Casas et al., 2005), there has been little empirical studies on trust in social cognition, let alone on its specific role in human communication and information exchange. The same is true for novelty.

Internal Models of Sources

Present-day individuals and organizations are often confronted with fundamental problems

created by an ever more complex and fast changing environment. Therefore, we must immediately know how to prioritize the incoming information and we have to decide which information is to be trusted and valid, and which is not. An individual or organization can efficiently manage this constant flow of information, when it memorizes trustworthy sources in case similar problems occur later.

Surprisingly, past research on group structure addressed predominantly pre-established or formal communication networks with fixed roles for individuals rather than flexible communication routes (Monge & Eisenberg, 1987). For instance, researchers explored the role and efficiency of information routing in centralized and hierarchical networks as opposed to more decentralized networks (Jablin, 1979, 1987; Leavitt, 1951; Mackenzie, 1976; Shaw, 1964, 1978). It was found that as soon as tasks become more complex and multifaceted, groups tend to gravitate naturally to more decentralized networks (Brown & Miller, 2000; Shaw, 1978). The question then is, how under these more dynamic structures, do people seek the most relevant and valid source so that available but dormant information is exploited more efficiently? Given the crucial importance of selecting trustworthy sources, can the TRUST model reproduce this important human skill? Yes, it can. We illustrate this with two brief simulations.

Simulation 7: Effective Communication Channels. The learning history of the simulations is shown in Table 9. The basic idea is that people should look for sources that share their background knowledge if they search for information on known topics, and should select sources having a different knowledge base when searching for novel information.

In simulation 7a, the concept of an "Expert" source is not yet explicitly incorporated in the representation of the "Seeking" agent. As can be seen form the top panel of the Test Phase, in this case trust is simply gleaned from recreating what the Expert agent said and testing how it fits with the Seeking agent's own beliefs. This testing procedure reactivates or uses the Seeker←Expert trust weights, which implement a limited internal model of the Expert. Although sufficient for a conversation, this does not allow the agent to build up a full-blown model on the trustworthiness of an Expert. To address this, in simulation 7b, the notion of an Expert agent is incorporated in the memory of the Seeking agent, so that associations can be built between the potential Expert and his or her knowledge on a known or novel topic. The testing procedure (in the bottom panel of the Test Phase) then simply consists of priming these memory associations.

Results. As can be seen in Figure 14, both simulations performed as predicted. The expert with the same background was preferred for information on known topics, while the expert with a different background was selected for information on novel issues. An ANOVA reveals a significant interaction with Background (same vs. novel) and Information Search (old vs. new) for the first simulation, $F(1, 196) = 157305$, $p < .001$, and the second simulation, $F(1, 196) = 30476$, $p < .001$.

Antecedents of Trust

How is trust in information provided by other agents increased or decreased? The TRUST model proposes that if no a priori expectations on agents exist, people trust information so long as it *fits with their own beliefs*. Although some degree of divergence is tolerated, if the discrepancy is too high, the information is not trusted and hence does not influence people's own belief system. Thus, rather than some internal inconsistency or some internal ambiguities in the story told, it is the inconsistency with one's own beliefs that sets off the listener and make him or her to distrust the information. These opposing predictions can be directly tested by comparing perceived trust in information that has low or high internal inconsistency versus low or high consistency with prior beliefs.

Consequences of Trust

In general, the TRUST model predicts that more trust results in more belief change and more adoption of collective views and solutions. There are several ways to explore this prediction. For instance, *directly* by asking participants of a discussion under conditions of trust or distrust, how much they thought the information provided by some agents was useful, how much they privately belief the consensus reached (when the task involves a collective decision) or how much they agree with the group solution (when the task is to find a solution to a problem). Or more *indirectly*, for instance, by measuring the time it took to reach a consensus or a solution in the group, or by the number of disagreement in the group, opposition or denigration of others, as well as other process variables. Under more controlled laboratory conditions, the consequences of trust can be measured by the response time in answering questions. The prediction is that it takes more time to read and understand untrustworthy information.

The Automaticity of Trust and Novelty

Our simulations suggest that trust is developed and applied automatically, outside of consciousness, rather than being a deliberate, controlled process. 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. To test that trust is automatically applied, one can adopt an experimental paradigm on spontaneous inferences (see e.g., Van Overwalle, Drenth & Marsman, 1999). For instance, one can compare messages communicated by trusted and distrusted sources. We predict that spontaneous inferences about people in these messages or made only when they are considered trustworthy, demonstrating that trust is automatically applied. For instance, when reading the sentence "Jaana solved the mystery halfway the book" the spontaneous inference that Jaana is intelligent is activated only when this message was provided by trusted sources. Similarly, using the same paradigm, we can explore to what extent speakers spontaneously facilitate the activation of novel story elements at the expense of known story elements when they have to communicate a message.

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. Specifically, acquired trust in the information provided by communicators was seen as an essential social and psychological requirement for any model of communication. This was implemented in the model on the basis of the consistency of the incoming information with the receiving agents' existing believes and past experiences. Trust leads to a selective filtering out of less reliable data and selective propagation of novel information, and so biases information transmission. From this implementation of 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 shared 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. These communicative aspects of our connectionist implementation were illustrated in a number of simulations of key communication phenomena, including attitude shifts in persuasive communication (Simulations 1 & 2), convergence in opinions and beliefs (Simulation 3), and propagation of biased information (Simulations 4—6).

Perhaps one of the major contributions of the model that makes this possible is its dynamic nature. 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 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. The model also incorporates a number of important criteria of human communication, such as the maxims of Grice (1975) as well as the capacity for each individual to have his or her own representation and perspective on reality.

Phenomena such as polarization, spreading of rumors, increasing stereotyping, and the failure to consider all relevant (unique) information or possibilities point us to the danger that at least under some circumstances, the processes of communicating information among the members of a group seems to make their collective cognition and judgments less reliable. The present paper helps us to illuminate and tear apart some basic mechanism in the creation of communication biases and misperceptions.

References

Adolphs, R., & Damasio, A. (2001). The interaction of affect and cognition: A neurobiological perspective. In J.P. Forgas (Ed.). Handbook of affect and social cognition (pp. 27-49). Mahwah, NJ: Lawrence Erlbaum Associates.

Ajzen, I. & Madden, T.J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. Journal of Experimental Social Psychology, 22, 453—474.

Ajzen, I. (1988). Attitudes, personality and behavior. Homewood, IL: Dorsey Press.

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50, 179—211.

Ajzen, I. (2002). Residual effects of past on later behavior: Habituation and reasoned action perspectives. Personality and Social Psychology Review, 6, 107—122.

Allan, L. G. (1993). Human contingency Judgments: Rule based or associative? *Psychological Bulletin, 114*, 435-448.

Anderson, N. H. (1967). Averaging model analysis of set size effect in impression formation. Journal of Experimental Psychology, 75, 158—165.

Anderson, N. H. (1971). Integration theory and attitude change. Psychological Review, 78, 171-206.

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

Ans, B. & Rousset, S. (2000). Neural networks with a self-refreshing memory: Knowledge transfer in sequential learning tasks without catastrophic forgetting. Connection Science, 12, 1-19.

Ans, B., & Rousset, S. (1997). Avoiding catastrophic forgetting by coupling two reverberating neural networks. Académie des Sciences de la vie, 320, 989-997.

Aunger R. (ed.) (2001). *Darwinizing culture: The status of semetics as a science*. Oxford, UK: Oxford University Press

Axsom, D. Yates, S. & Chaiken, S. (1987). Audience response as a heuristic cue in persuasion. Journal of Personality and Social Psychology, 53, 30—40.

Baker, A. G., Berbier, M. W., & Vallée-Tourangeau, F. (1989). Judgments of a 2 x 2 contingency table: sequential processing and the learning curve. The Quarterly Journal of Experimental Psychology, 41B, 65—97.

Baker, A. G., Mercier, P., Vallée-Tourangeau, F., Frank, R. & Pan, M. (1993). Selective associations and causality judgments: Presence of a strong causal factor may reduce judgments of a weaker one. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*, 414-432.

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. Journal of Personality and Social Psychology, 87, 5—22.

Bargh, J. A., Chaiken, S., Govender, R. & Pratto, F. (1992). The generality of the automatic attitude activation effect. Journal of Personality and Social Psychology, 62, 893—912.

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

Berkowitz, L. & Knurek, D. a. (1969). Label-mediated hostility generalization. Journal of Personality and Social Psychology, 13, 200—206.

Betsch, T., Plessner, H., & Schallies, E. (2004). The value-account model of attitude formation. In G. R. Maio, G. Haddock (Eds.) *Theoretical perspectives on attitudes for the 21st century - The Gregynog Symposium*. Psychology Press, in press.

Betsch, T., Plessner, H., Schwieren, C., & Gütig, R. (2001). I like it but I don't know why: A value-account approach to implicit attitude formation. *Personality and Social Psychology Bulletin, 27*, 242—253.

Bohner, G. & Weinerth, T. (2001). Negative affect can increase or decrease message scrutiny: the affect interpretation hypothesis. *Personality and Social Psychology Bulletin, 27*, 1417—1428.

Bohner, G., Ruder, M., & Erb, H.-P. (2002). When expertise backfires: Contrast and assimilation effects in persuasion. *British Journal of Social Psychology, 41*, 495—519.

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

Bower, G. H. (1981) Emotional mood and memory. *American Psychologist, 36*, 129—148.

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.

Brown, T. M. & Miller, C. E. (2000). Communication networks in task-performing groups: Effects of task complexity, time pressure, and interpersonal dominance. *Small Group Research, 31,* 131-157.

Bura, S., Guérin-Pace, F., Mathian, H., Pumain, D., & Sanders, L. (1995). Cities can be agents too: A model for the evolution of settlement systems. In G. N. Gilbert & R. Conte (Eds.) *Artificial societies: The computer simulation of social life* (pp. 86—103) London, UK: UCL Press.

Canli, T., Desmond, J. E., Zhao, Z., Glover, G. & Gabrieli, J. D. E. (1998). Hemispheric asymmetry for emotional stimuli detected with fMRI. NeuroReport, 9, 3233—3239.

Chaiken, S. & Eagly, A. H. (1983). Communication modality as a determinant of persuasion: The role of communicator salience. Journal of Personality and Social Psychology, 45, 241—256.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. Journal of Personality and Social Psychology, 39, 752—766.

Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.). Social influence: The Ontario Symposium (Vol. 5., pp. 3—39). Hillsdale, NJ: Erlbaum.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment. Journal of Personality and Social Psychology, 66, 460—473.

Chaiken, S., Duckworth, K. L., & Darke, P. (1999). When parsimony fails… *Psychological Inquiry, 10*, 118—123.

Chaiken, S., Liberman, A. & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.) *Unintended thought* (pp. 212—252). New York, NY: Guilford.

Chapman, G. B. & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory and Cognition, 18*, 537-545.

Chapman, G. B. & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory and Cognition, 18*, 537—545.

Chapman, G. B. (1991). Trial order affects cue interaction in contingency judgment *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 837-854.

Chen, S. & Chaiken, S. (1999). The Heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.). *Dual-process theories in social psychology* (pp. 73—96). New York, NY: Guilford Press.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*, 545-567.

Chi, M. T. H. (1989). Assimilating evidence: The key to revision? *Behavioral and Brain Sciences, 12*, 470-471.

Cooper, J. & Fazio, R. H. (1984). A new look at dissonance theory. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 17, pp. 229-266). New York: Academic Press.

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.

Cunningham, W. A., Johnson, M. K., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2003). Neural components of social evaluation. *Journal of Personality and Social Psychology, 85*, 639—649.

Darke, P. R., Chaiken, S., Bohner, G., Einwiller, S., Erb, H-P., Hazlewood, J. D. (1998). Accuracy motivation, consensus information, and the low of large numbers: Effects on attitude judgment in the absence of argumentation. *Personality and Social Psychology Bulletin, 24*, 1205—1215.

De Houwer, K., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853—869.

Dijksterhuis, A. (2002, September). *A subliminal road to happiness: Enhancing implicit self-esteem by subliminal evaluative conditioning*. Paper presented at the 4th European Social Cognition Network Meeting, Paris, France.

Eagly, A. H. & Chaiken, S. (1993). *The psychology of Attitudes.* San Diego, CA: Harcourt Brace.

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.

Eiser, J. R., Fazio, R. H., Stafford, T., & Prescott, T. J. (2003). Connectionist Simulation of Attitude Learning: Asymmetries in the Acquisition of Positive and Negative Evaluations. *Personality and Social Psychology Bulletin, 29*, 1221—1235

Erb, H.-P., Bohner, G., Schmälzle, K., & Rank, S. (1998). Beyond conflict and discrepancy: Cognitive bias in minority and majority influence. *Personality and Social Psychology Bulletin, 24*, 620-633.

Estes, W. K. (1994) *Classification and cognition.* New York, NY: Oxford University Press.

Estes, W. K., Campbell, J. A., Hatsopoulos N. & Hurwitz, J. B. (1989). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology, Learning, Memory and Cognition, 15*, 556-571.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: the MODE model as an integrative framework. In M. P. Zanna (Ed.) *Advances in Experimental Social Psychology* (vol. 13, pp. 75—109). San Diego, CA: Academic Press.

Fazio, R. H., Powell, M. C. (1997). On the value of knowing one's likes and dislikes: Attitude accessibility, stress, and health in college. *Psychological Science, 8*, 430—436.

Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C., & Kardes, F.R. (1986). On the automatic activation of

attitudes. *Journal of Personality and Social Psychology*, *50*, 229-238.
Ferber, J. (1989) *Des objects aux agents*. Doctoral thesis, University of Paris VI, France.
Festinger, L. (1957) *A theory of cognitive dissonance.* Evanston, IL: Row, Peterson.
Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environment. *Journal of Personality and Social Psychology, 103*, 193-214.
Fiedler, K., Walther, E. & Nickel, S. (1999). The auto-verification of social hypotheses: Stereotyping and the power of sample size. *Journal of Personality and Social Psychology, 77*, 5-18.
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.
Fiske, S. F. (2005). Social beings: A core motives approach to social psychology. Hoboken, JN: Wiley.
Forgas, J. P. (2001) *Handbook of affect and social cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.
Försterling, F. (1989). Models of covariation and attribution: How do they relate to the analogy of analysis of variance? *Journal of Personality and Social Psychology, 57*, 615—625.
French, R. (1997). Pseudo-recurrent connectionist networks: An approach to the "sensitivity–stability" dilemma. *Connection Science, 9*, 353-379.
French, R. (1999). Catastrophic forgetting in connectionist networks: Causes, consequences and solutions. *Trends in Cognitive Sciences, 3*, 128—135.
Gabrys, G. & Lesgold, A. (1989). Coherence: Beyond constraint satisfaction. *Behavioral and Brain Sciences, 12*, 475.
Gluck, M. A. & Bower, G. H. (1988a). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*, 227-247.
Gluck, M. A. & Bower, G. H. (1988b). Evaluating an adaptive network model of human learning. *Journal of Memory and Language, 27*, 166-195.
Gollwitzer, P. M. (1990) Action phases and mind-sets. In E. T. Higgins and R. M. Sorrentino (Eds.), Handbook of motivation and cognition: Foundations of social behavior (Vol. 2, pp. 53—92). New York: Guilford Press.
Greenwald, A. G. (1968). Cognitive learning, cognitive response to persuasion, and attitude change. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes* (pp.147—170). San Diego, CA: Academic Press.
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.
Haddock, G, Rothman, A. J., & Schwarz, N. (1996). Are (some) reports of attitude strength context dependent? *Canadian Journal of Behavioral Science, 28*, 313—316.
Haddock, G. (2000). Subjective ease of retrieval and attitude-relevant judgments. In H. Bless & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior* (pp. 125-142). Philadelphia: Taylor & Francis.
Hamilton, D. L., Driscoll, D. M. & Worth, L. T. (1989). Cognitive organization of impressions: Effects of incongruency in complex representations. *Journal of Personality and Social Psychology, 57*, 925-939.
Hansen, D. H. & Hall, C. A. (1985). Discounting and augmenting facilitative and inhibitory forces: the winner takes almost all. *Journal of Personality and Social Psychology, 49*, 1482-1493.
Hastie, R. (1980). Memory for behavioral information that confirms or contradicts a personality impression. In R. Hastie, T. M. Ostrom, E. B. Ebbesen, R. S. Wyer, D. L. Hamilton, & D. E. Carlston (Eds.). *Person Memory: The cognitive basis of social perception* (pp. 155—177). Hillsdale, NJ: Erlbaum.
Haugtvedt, C. P., Schumann, D. W., Schneier, W. L., Warren, W. L. (1994) Advertising repetition and variation strategies: Implications for understanding attitude strength. *Journal of Consumer Research, 21*, 176—189
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.
Heesacker, M., Petty, R.E., Cacioppo, J. T. (1983). Field dependence and attitude change: Source credibility can alter persuasion by affecting message-relevant thinking. *Journal of Personality, 51*, 653—666.
Heylighen F. (1998). What makes a meme successful? Selection criteria for cultural Evolution. *Proceeding of the 16th International Congress on Cybernetics* (pp. 423-418). Association International de Cybernétique, Namur.
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.
Hilton, D. J., Smith, R. H., & Kim S. H. (1995). The process of causal explanation and dispositional attribution. *Journal of Personality and Social Psychology, 68*, 377-387.
Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In J. L.

McClelland & D. E. Rumelhart (Eds.) *Parallel distributed processing. Explorations in the microstructure of cognition: Foundations* (vol. 1, pp. 77-109). Cambridge, MA: Bradford Books.

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 life* (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.

Isen, A.M. (1984). Towards understanding the role of affect in cognition. In R.S. Wyer, Jr. & T.K. Srull (Eds.), *Handbook of social cognition* (Vol. 3, pp. 179-236). Hillsdale, NJ: Erlbaum.

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

Ito, T.A., & Cacioppo, J.T. (2001). Affect and attitudes: A social neuroscience approach. In J.P. Forgas (Ed.) *Handbook of affect and social cognition* (pp. 50-74). Mahwah, NJ: Lawrence Erlbaum Associates.

Jablin, F. M. (1979). Superior-Subordinate communication: The state of the art. *Psychological Bulletin, 86*, 1201—1222.

Jablin, F. M. (1987). Formal Organizational Structure. In F. M. Jablin, L. L. Putnam, & K. H. Roberts (Eds.) *Handbook of organizational communication: an interdisciplinary perspective* (pp. 389—419). London, UK: Sage.

Janis, I. L. (1972). *Victims of groupthink*. Boston: Houghton Mifflin.

Jonas, K., Diehl, M. & Brömer, P. (1997). Effects of attitudinal ambivalence on information processing and attitude-intention consistency. *Journal of Experimental Social Psychology, 33*, 190—210.

Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement: A suggested modification of the semantic differential technique. *Psychological Review, 77*, 361—372.

Kashima, Y, & Kerekes, A. R. Z. (1994). A distributed memory model of averaging phenomena in person impression formation. *Journal of Experimental Social Psychology, 30*, 407—455.

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.

Katz, D., & Stotland, E. (1959). A preliminary statement to a theory of attitude structure and change. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3, pp. 423-475). New York: McGraw-Hill.

Kelley, H. H. (1967). Attribution in social psychology. *Nebraska Symposium on Motivation, 15*, 192-238.

Kelley, H. H. (1971). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins & B. Weiner (Eds.) *Attribution: Perceiving the causes of behavior* (pp. 1—26). Morristown, NJ: General Learning Press.

Kelley, H. H. (1971a). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins & B. Weiner (Eds.) *Attribution: Perceiving the causes of behavior* (pp. 1-26). Morristown, NJ: General Learning Press.

Kelley, H. H. (1971b). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins & B. Weiner (Eds.) *Attribution: Perceiving the causes of behavior* (pp. 151-174). Morristown, NJ: General Learning Press.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R. & Montague, R. (2005). Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange. *Science, 308*, 78—83.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R. & Montague, R. (2005). Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange. *Science, 308*, 78—83.

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.

Kruglanski, A. W. & Thompson, E. P. (1999). Persuasion by a single route: A view from the unimodel. *Psychological Inquiry, 10*, 10, 83—109.

Kruglanski, A. W., Schwartz, S. M., Maides, S., & Hamel, I. Z. (1978). Covariation, discounting, and

augmentation: Towards a clarification of attributional principles. *Journal of Personality, 76,* 176–189.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1083—1119.

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

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

LaBerge, D. (1997). Attention, awareness and the triangular circuit. *Consciousness and Cognition, 6*, 149—181.

LaBerge, D. (2000). Networks of attention. In M. S. Gazzaniga (Ed.), *The New Cognitive Neuroscience* (pp. 711—724). Cambridge: MIT Press.

Lane, R. D., Reiman, E. M. , Bradley, M. M., Lang, P. J., Ahern, G. L., Davidson, R. J. & Schwartz, G. E. (1997). Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia, 35*, 1437—1444.

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.

Leavitt, H. J. (1951). Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology, 46*, 38—50.

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

Lieberman, M. D., Ochsner, K. N., Gilbert, D. T., & Schacter, D. L. (2001). Attitude change in amnesia and under cognitive load. *Psychological Science, 12*, 135—140.

Linder, D.E., Cooper, J. & Jones, E.E. (1967). Decision freedom as a determinant of the role of incentive magnitude in attitude change. Journal of Personality and Social Psychology, 6, 245-254.

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.

Mackenzie, K. D. (1976). *A theory of group structure*. New York: Gordon & Breach.

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

Mackie, D. M. & Worth, L. T. (1989). Processing deficits and the mediation of positive affect in persuasion. *Journal of Personality and Social Psychology, 57*, 27—40.

Mackie, D. M. (1987). Systematic and non-systematic processing of majority and minority persuasive communications. *Journal of Personality and Social Psychology, 53*, 41—52.

Maheswaran, D. & Chaiken, S. (1991). Promoting systematic processing in low-motivation settings: Effect of incongruent information on processing and judgment. *Journal of Personality and Social Psychology, 61*, 13—25.

Manis, M., Dovalina, I., Avis, N. E., & Cardoze, S. (1980). Base rates can affect individual predictions. *Journal of Personality and Social Psychology, 38*, 231—248.

Marr, D. (1982). Vision. San Francisco, CA: Freeman.

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. (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.

McClelland, J. L., McNaughton, B., & O'Reilly, R. (1995). Why there are complementary learning systems

in the hippocampus and neocortex: Insights from the successes and the failures of connectionist models of learning and memory. *Psychological Review, 102*, 419-457.

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

McCloskey, M., & Cohen N.J. (1989). Catastrophic interference in connectionist networks: the sequential learning problem. *The Psychology of Learning and Motivation, 24*, 109-165.

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

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

Miller, N. & Colman, D. E. (1981). Methodological issue in analyzing the cognitive mediation of persuasion. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion* (pp. 105—125). Hillsdale, NJ: Erlbaum.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1986). Assessment of the Rescorla-Wagner model. *Psychological Bulletin, 117*, 363-386.

Monge, P. E., & Eisenberg, E. M. (1987). Emergent communication networks. In F. M. Jablin, L. L. Putnam, & K. H. Roberts (Eds.) Handbook *of organizational communication: an interdisciplinary perspective* (pp. 304—342). London, UK: Sage.

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.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.

Ochsner, K. N. & Lieberman, M. D. (2001). The emergence of social cognitive neuroscience. *American Psychologist, 56*, 717—734.

Olson, M. A. & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413—417.

O'Reilly, R. C. & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review, 108*, 311—345.

Pacton, S., Perruchet, P., Fayol, M. & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. Journal of Experimental Psychology: General, 130, 401—426.

Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review, 101*, 587-607.

Penrod, S. & Hastie, R. (1980). A computer simulation of jury decision making. *Psychological Review, 87*, 133-159.

Petty, R. E. & Cacioppo, J. T. (1981). *Attitudes and persuasion: Central and peripheral routes to attitude change*. New York: Springer.

Petty, R. E. & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology, 46*, 69—81.

Petty, R. E. & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 19, pp. 123—205). San Diego, CA: Academic Press.

Petty, R. E. & Wegener, D. T. (1999). The elaboration likelihood model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.). *Dual-process theories in social psychology* (pp. 41—72). New York, NY: Guilford Press.

Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *Journal of Consumer Research, 10*, 135—147.

Petty, R. E., Schumann, D. W., Richman, S. A., & Strathman, A. J. (1993). Positive mood and persuasion: Different roles for affect under high- and low-elaboration conditions. *Journal of Personality and Social Psychology, 64*, 5—20.

Petty, R.E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology, 41*, 847—855.

Posner, M. I. (1992). Attention as a cognitive and neural system. *Current Directions in Psychological Science, 1*, 11-14.

Powell, M. C., & Fazio, R. H. (1984). Attitude accessibility as a function of repeated attitudinal expression. *Personality and Social Psychology Bulletin, 10*, 139—148.

Priester, J. R. & Petty, R. E. (1996). The gradual threshold model of ambivalence: Relating the positive and negative bases of attitudes to subjective ambivalence. *Journal of Personality and Social*

Psychology, 71, 431—449.
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.
Ratcliff, R. (1990). Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychological Review, 97*, 285-308.
Ratneshwar, S. & Chaiken, S. (1991). Comprehension's role in persuasion: the case of its moderating effect on the persuasive impact of source cues. *Journal of Consumer Research, 18*, 52—62.
Read, S. J. & Marcus-Newhall, A. (1993) Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology, 65*, 429-447.
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. (1997) *Connectionist models of Social Reasoning and Social Behavior*. Lawrence Erlbaum, in press.
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.
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.
Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement (pp. 64-98). In A. H. Black & W. F. Prokasy (Eds.) *Classical conditioning II: Current research and theory.* New York, NY: Appleton.
Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.) Classical *conditioning II: Current research and theory* (pp. 64–98). New York: Appleton-Century-Crofts.
Riketta, M., & Dauenheimer, D. (2002). *Manipulating self-esteem through subliminally presented words*. Manuscript submitted for publication.
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.
Romero, A. A., Agnew, C. R., & Insko, C. A. (1996). The cognitive mediation hypothesis revisited: An empirical response to methodological and theoretical criticism. *Personality and Social Psychology Bulletin, 22,* 651-665.
Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89-115.
Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89-115.
Rosenberg, M. J. & Hovland, C. I. (1960). Cognitive, affective and behavioral components of attitudes. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1-14). Hew Haven, CT: Yale University Press.
Roskos-Ewoldsen, D. R., Bichsel, J. & Hoffman, K. (2002). The influence of accessibility of source likeability on persuasion. *Journal of Experimental Social Psychology, 38*, 137—143.
Rumelhart, D. & McClellan, J. (1986). ON learning the past tense of English verbs: Implicit rules or parallel distributed processing? In J. M. McClelland & D. E. Rumelhart (1986). *Parallel distributed processing* (Vol. 2, pp. 216—271). Cambridge, MA: Bradford.
Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In D. E. Rumelhart & J. L. McClelland (Eds.) *Parallel distributed processing. Explorations in the microstructure of cognition: Psychological and biological models* (vol. 2, pp. 7-57). Cambridge, MA: Bradford Books.
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.
Schumann, D. W., Petty, R., & Clemons, D. S. (1990). Predicting the effectiveness of different strategies of advertising variation. A test of the repetition-variation hypothesis. *Journal of Consumer Research, 17*, 192—201.

Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In E.T. Higgins & R. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2). New York: Guilford.

Schwarz, N., & Clore G.L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*, 513-523.

Schwarz, N., Bless, H., & Bohner, G. (1991). Mood and persuasion: Affective states influence the processing of persuasive communications. *Advances in Experimental Social Psychology, 24*, 161-199.

Senge P. (1990). *The Fifth Discipline: the Art and Practice of Learning Organizations.* Doubleday.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology, 37b*, 1-21.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology, 37b*, 1—21.

Shanks, D. R. (1987). Acquisition functions in contingency judgment. *Learning and Motivation, 18*, 147—166.

Shanks, D. R. (1991a). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 433-443.

Shanks, D. R. (1991b). On similarities between causal judgments in experienced and described situations. *Psychological Science, 2*, 341-350.

Shanks, D. R. (1993). Human instrumental learning: A critical review of data and theory. *British Journal of Psychology, 84*, 319-354.

Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology, 48a*, 257-279.

Shanks, D.R., Lopez, F. J., Darby, R. J., Dickinson, A. (1996). Distinguishing associative and probabilistic contrast theories of human contingency judgment. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.) *The psychology of learning and motivation* (Vol. 34, pp. 265—311). New York, NY: Academic Press.

Shaw, M. E. (1964). Communication networks. In L. Berkowitz (Ed.) *Advances in Experimental Social Psychology, 1*, 111-147. New York: Academic Press.

Shaw, M. E. (1978). Communication networks fourteen years later. In Berkowitz, L. (Ed.). *Group processes* (pp ??). New York: Academic Press.

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.

Shultz, T. R. & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological Review, 103*, 219-240.

Siebler, F. (2002). *Connectionist modeling of social judgment processes*. Unpublished PhD Thesis, University of Kent, Canterbury, United Kingdom.

Sinclair, R. C., Mark, M. M. & Clore, G. L. (1994). Mood-related persuasion depends on (mis)attributions. *Social Cognition, 12*, 309—326.

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. & 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. & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108-131.

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

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

Smith, E. R., & DeCoster, J. (1997). Person perception and stereotyping: Simulation using distributed representations in a recurrent connectionist network. In S. J. Read & L. C. Miller (Eds.) *Connectionist models of Social Reasoning and Social Behavior*. Lawrence Erlbaum, in press.

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. & 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.

Staats, A. W. & Staats, C. K. (1958). Attitudes established by classical conditioning. *Journal of Abnormal and Social Psychology, 57*, 37—40.

Stangor, C. & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin, 111*, 42—61.

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.

Strack, F. & Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review, 8*, 220-247.

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

Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.

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.

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.

Tormala, Z. L., Petty, R. E., Briñol, P. (2002). Ease of retrieval effect in persuasion: A self-validation analysis. *Personality and Social Psychology Bulletin, 28*, 1700—1712.

Van Duynslaeger, M. & Van Overwalle, F. (2004) Do Persuasive Heuristics Need Abstract Rules? Unpublished data, Free University of Brussels, Belgium.

Van Hamme & Wasserman (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation, 25*, 127-151.

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. (1996). The relationship between the Rescorla-Wagner associative model and the probabilistic joint model of causality. *Psychologica Belgica*, 36, 171-192.

Van Overwalle, F. (1997). Dispositional Attributions require the Joint application of the Methods of Difference and Agreement. *Personality and Social Psychology Bulletin*, 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 Overwalle, F., & Labiouse, C. (2004) A recurrent connectionist model of person impression formation. Personality and Social Psychology Review, 8, 28—61.

Van Overwalle, F., & Van Rooy, D. (1997). A Connectionist Approach to Causal Attribution. In S. J. Read & L. C. (Eds.) *Connectionist models of Social Reasoning and Social Behavior*. Lawrence Erlbaum, in press.

Van Overwalle, F., & Van Rooy, D. (1998) A Connectionist Approach to Causal Attribution. In S. J. Read & L. C. Miller (Eds.) *Connectionist and PDP models of Social Reasoning and Social Behavior* (pp. 143—171). Lawrence Erlbaum

Van Overwalle, F., Heylighen, F. & Heath, M. (2005) Trust *in Communication between Individuals: A Connectionist Approach.* Proceedings of the Cognitive Science 2005 Workshop: Toward Social Mechanisms of Android Science.

Van Overwalle, F., Heylighen, F. & Heath, M. (2005) Trust *in Communication between Individuals: A Connectionist Approach.* Proceedings of the Cognitive Science 2005 Workshop: Toward Social Mechanisms of Android Science.

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.

Van Rooy, D., Van Overwalle, F., Vanhoomissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review, 8*, 28—61.

Wänke, M. & Bless, H. (2000). The effects of subjective ease of retrieval on attitudinal judgments: The moderating role of processing motivation. In H. Bless & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior* (pp. 143-161). Philadelphia: Taylor & Francis.

Wänke, M., Bless, H., & Biller, B. (1996). Subjective experience versus content of information in the construction of attitude judgments. *Personality and Social Psychology Bulletin, 22*, 1105-1113.

Wänke, M., Bohner, G., & Jurkowitsch, A. (1997). There are many reasons to drive a BMW: Does imagined

ease of argument generation influence attitudes? *Journal of Consumer Research, 24*, 170-177.
Wegener, D. T. & Petty, R. E. (1996). Effects of mood on persuasion processes: Enhancing, reducing and biasing scrutiny of attitude-relevant information. In L. L. Martin & A. Tesser. *Striving and Feeling: Interactions among goals, affect and self-regulation (pp. 329—362)*.
Weiner, B. (1985a). "Spontaneous" causal thinking. *Psychological Bulletin, 97*, 74-84.
Wilkes-Gibbs, D. & Clark, H. H. (1992). Coordinating beliefs in conversation. Journal of Memory and Language, 31, 183—195.
Wilson, T. D., Lindsey, S. & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101—126.
Wittenbaum, G. M. & Bowman, J. M. (2003). A social validation explanation for mutual enhancement. *Journal of Experimental Social Psychology,* 40, 169—184.
Wood, W., Kallgren, C. A., Preisler, R. M. (1985). Access to attitude-relevant information in memory as a determinant of persuasion: The role of message attributes. *Journal of Experimental Social Psychology, 21*, 73—85.
Worth, L. T. & Mackie, D. M. (1987). Cognitive mediation of positive affect in persuasion. *Social Cognition, 5*, 76—94.
Zanna, M.P., Kiesler, C.A., & Pilkonis, P.A. (1970). Positive and negative attitudinal affect established by classical conditioning. *Journal of Personality and Social Psychology*, *14*, 321-328.

Footnotes

Table 1

Summary of the Main Simulation Steps in the TRUST Model

Step / Cycle	Equation in Text
A. setting external activation within all agents	*(1)*
B. activation spreading within all non-listening agent	*(2) (3) (4b)*
C. attenuation and boosting of internally generated and expressed activation (see "i" in the tables 4 — 8)	*(7a) (7b)*
D. spreading of activation from talking to listening agents (see "i" and "?" respectively in the tables 4 — 8)	*(6)*
E. activation spreading within listening agents	*(2) (3) (4b)*
F. trust weight update (between agents)	*(8)*
G. connection weight update (within agents)	*(5)*

Note. "Within" refers to all units within an agent.

Table 2

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 = ± .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

Table 3

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]
2	Trust in Source	More opinion shift if trust in the source is high	More information transmission if trust weight is high [a]
i	Polarization	More opinion shift after group discussion	More information transmission by a majority
		Communication and Stereotyping	
3	Referencing	Less talking is needed to identify objects	Overactivation of talker's and listener's net
ii	Word use	Acquiring word terms, synonyms and ambiguous words	Information transmission on word meaning and competition between word meanings [a]
4	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]
5	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
6	Sharing Unique Information	Unique information is communicated only after some time in a free discussion	Same as Simulation 5
		Communication Channels	
7	Trust in Agents	• How to detect trustworthy channels • How to built knowledge about this	• Test receiving trust weights • Built associations with agent unit

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

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). Cell entries denote external activation and empty cell denote 0 activation; #=frequency of trial; i=internal activation (the talking agent generates this by activating the topical issue) is taken as external activation; ? = (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

Persuasive Arguments and Trust in the Source (Simulation 2)

	Talking Agent					Listening Agent				
	Topic	Feat1	Feat2	Feat3	Feat4	Topic	Feat1	Feat2	Feat3	Feat4
Setting Agent ♦ Listener trust weights to +1 for ingroup 0 for outgroup										
Prior Learning of Pro (*Anti*) Arguments										
#10	1	1 *(-1)*	1 *(-1)*	1 *(-1)*	1 *(-1)*					
Talking and Listening										
#5	1	i	i	i	i	?	?	?	?	?
Test of Beliefs										
of Listener						1	?	?	?	?

Note. Schematic version of learning experiences along Mackie & Cooper (1984). Cell entries denote external activation and empty cell denote 0 activation; #=frequency of trial; i=internal activation (the talking agent generates this by activating the topical issue) is taken as external activation; ? = (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 6

Referencing (Simulation 3)

	Talking Agent					Listening Agent				
	Picture Martini	Glass	Legs	Each-Side		Picture Martini	Glass	Legs	Each-Side	
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"							?	?	?	?
Test of Accuracy										
of "Matcher"						1	?	?	?	?

Note. Schematic version of learning experiences along Schober & Clark (1989, Experiment 1). Cell entries denote external activation and empty cell denote 0 activation; #=frequency of trial; i=internal activation (the talking agent generates this by activating the topical issue) is taken as external activation;? = (little ear) external activation received from the talking agent, in the test phase ? = activation of talking agents during the previous talking phase (test of talking) or the activation read off after priming the topic (test of accuracy). Trial order was randomized in each phase and condition.

Table 7*: Rumor Paradigm (Simulation 4)*

	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). Cell entries denote external activation and empty cell denote 0 activation; SC=Stereotype Consistent; SI=Stereotype Inconsistent; #=frequency of trial; i=internal activation (the talking agent generates this by activating the topical issue) is taken as external activation;? = (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 8

Shared vs. Unique Information (Simulation 6)

	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). Cell entries denote external activation and empty cell denote 0 activation; Uni=Unique; #=frequency of trial; i=internal activation (the talking agent generates this by activating the topical issue) is taken as external activation;? = (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 9

Trust and Communication Channels (Simulation 7)

		Seeking Agent 1					Expert Agent 2				
	[Agent2]	Topic	Known1	Known2	New1	New2	Topic	Known1	Known2	New1	New2
Learning by Seeking Agent & Learning the Same or Novel Background by Expert Agent											
#10		1	1	1							
#10 Same							1	1	1		
#10 Novel						1			1	1	
Listening to Expert Agent											
#10	*[1]*	1	?	?	?	?	1	i	i	i	i
7a: Test of Expert Communication Channel											
Trust for Old		?	?	?			1	1	1		
for New		?			?	?	1			1	1
7b: Test of Expert Agent											
Trust for Old	*[1]*	*[1]*	?	?							
for New	*[1]*	*[1]*			?	?					

Note. Schematic version of learning experiences in determining trustworthy sources. Cell entries denote external activation and empty cell denote 0 activation; #=frequency of trial; i=internal activation (the talking agent generates this by activating the topical issue) is taken as external activation;? = (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. Either the same or novel background information was learned by Agent 2. The first Learning Phase was used in both simulations, while the second Testing Phase differs between Simulation 7a and 7b. The external activation denoted between straight parentheses is for Simulation 7b only.

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. A generic auto-associator connectionist model of an individual agent (with 4 units).

Figure 3. Graphical illustration of learning by the delta algorithm. With 100% co-occurrence, the object always co-occurs with its feature converging to a connection weight of +1; with 50% co-occurrence, the system converges to a connection weight of +.50.

Figure 4. The functional role of trust weights in communication between agents.

Figure 5. 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 6. Simulation 2: Attitude Shifts in function of Group Source. Human data are denoted by bars, simulated values by broken lines. The human data are from Table 2 in "Attitude polarization: Effects of group membership" by D. Mackie & J. Cooper, 1984, *Journal of Personality and Social Psychology, 46*, p. 579. Copyright 1984 by the American Psychological Association.

Figure 7. Interlude Simulation i: Polarization in function of the Progress in the Discussion

Figure 8. Simulation 3a: [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 9. Simulation 3b: Accuracy in Identifying the Reference by the Matcher, Early and Late

Overhearer. 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. 218. Copyright 1989 by Academic Press.

Figure 10. Interlude Simulation ii: Lexical Acquisition for new, matched, synonymous and ambiguous words.

Figure 11. Simulation 4: 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 12. Simulation 5: 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 13. Simulation 6: 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 14. Simulation 7: [Top] Trust in Effective Communication Channel and [Bottom] Trust in Agents.

Figure 1

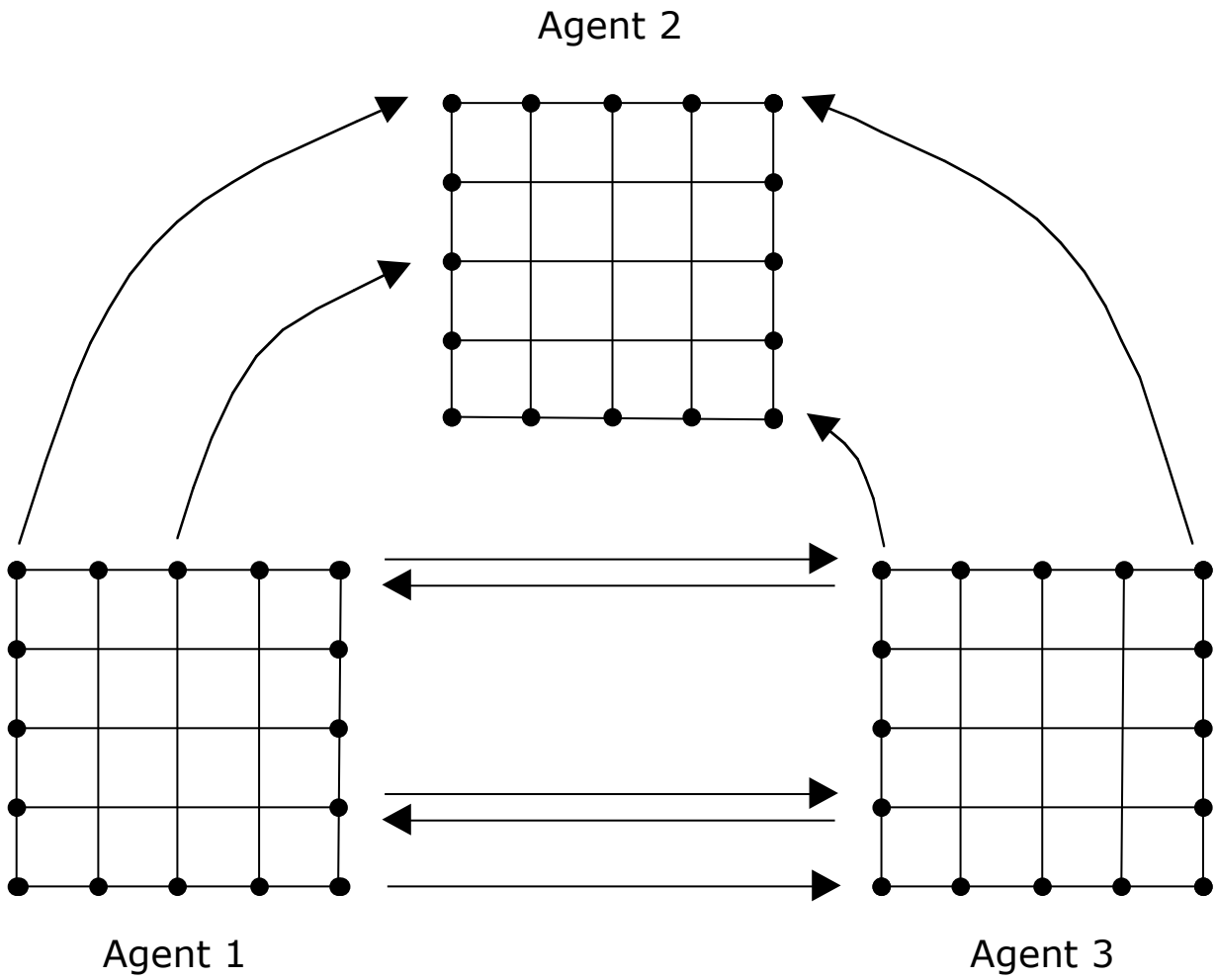


Figure 2

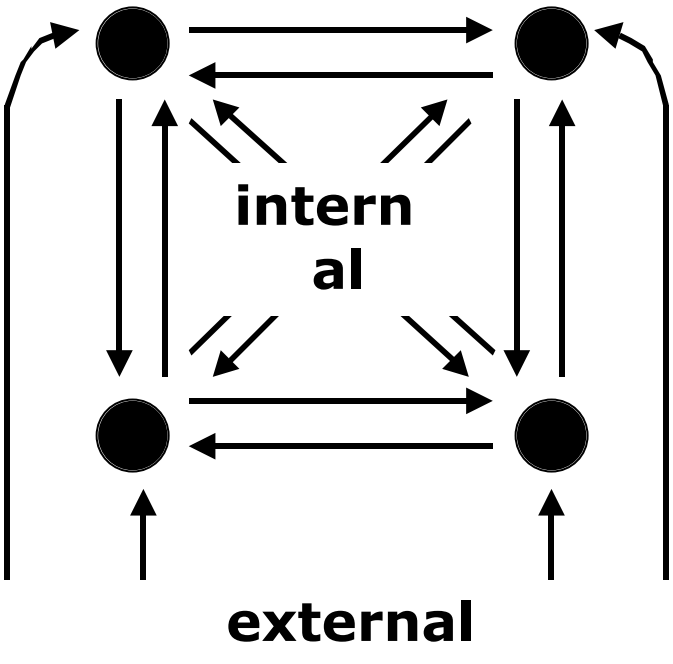


Figure 3

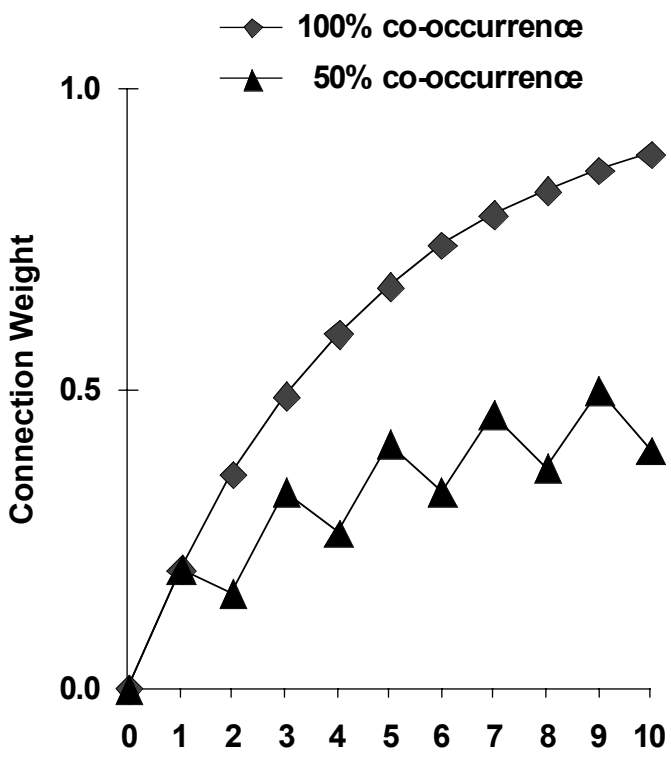

100% co-occurrence
50% co-occurrence
Connection Weight
1.0
0.5
0.0
0 1 2 3 4 5 6 7 8 9 10

Figure 4

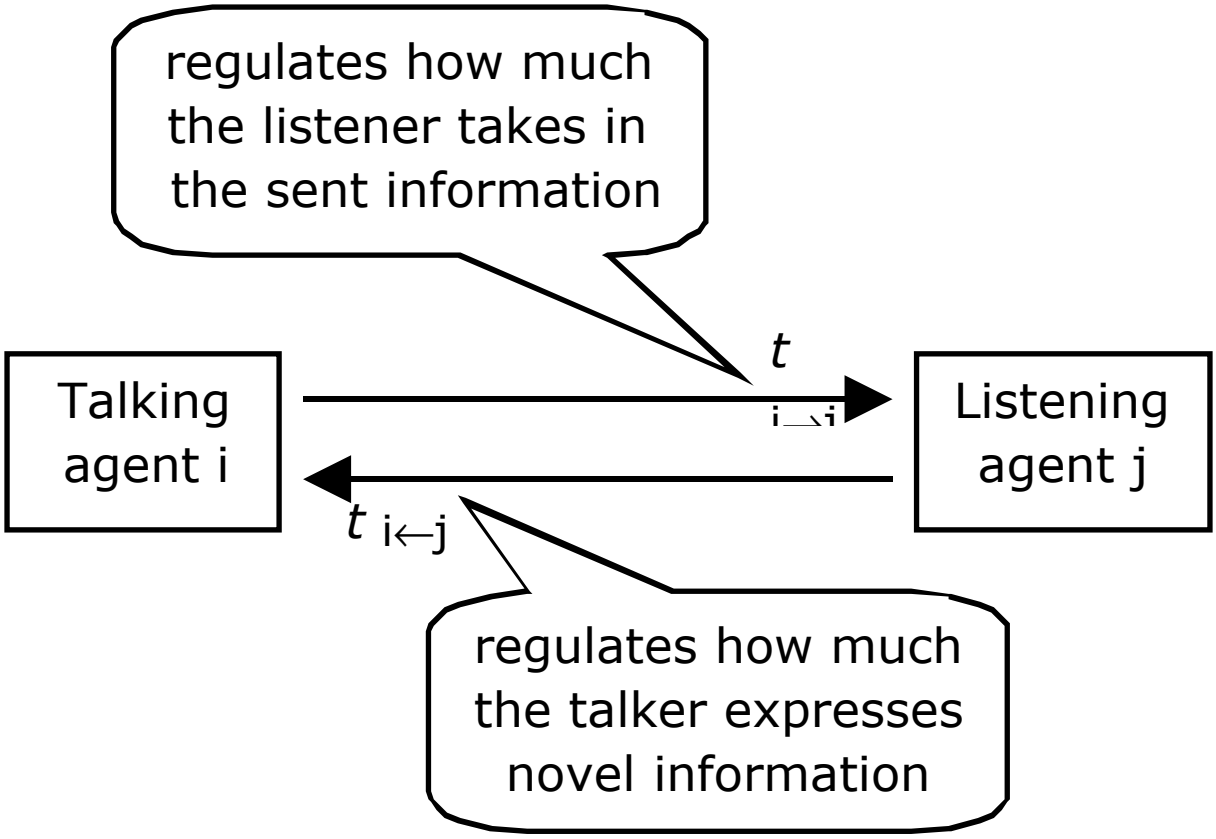

regulates how much the listener takes in the sent information
t
i→j
Talking agent i
Listening agent j
t i←j
regulates how much the talker expresses novel information

Figure 5

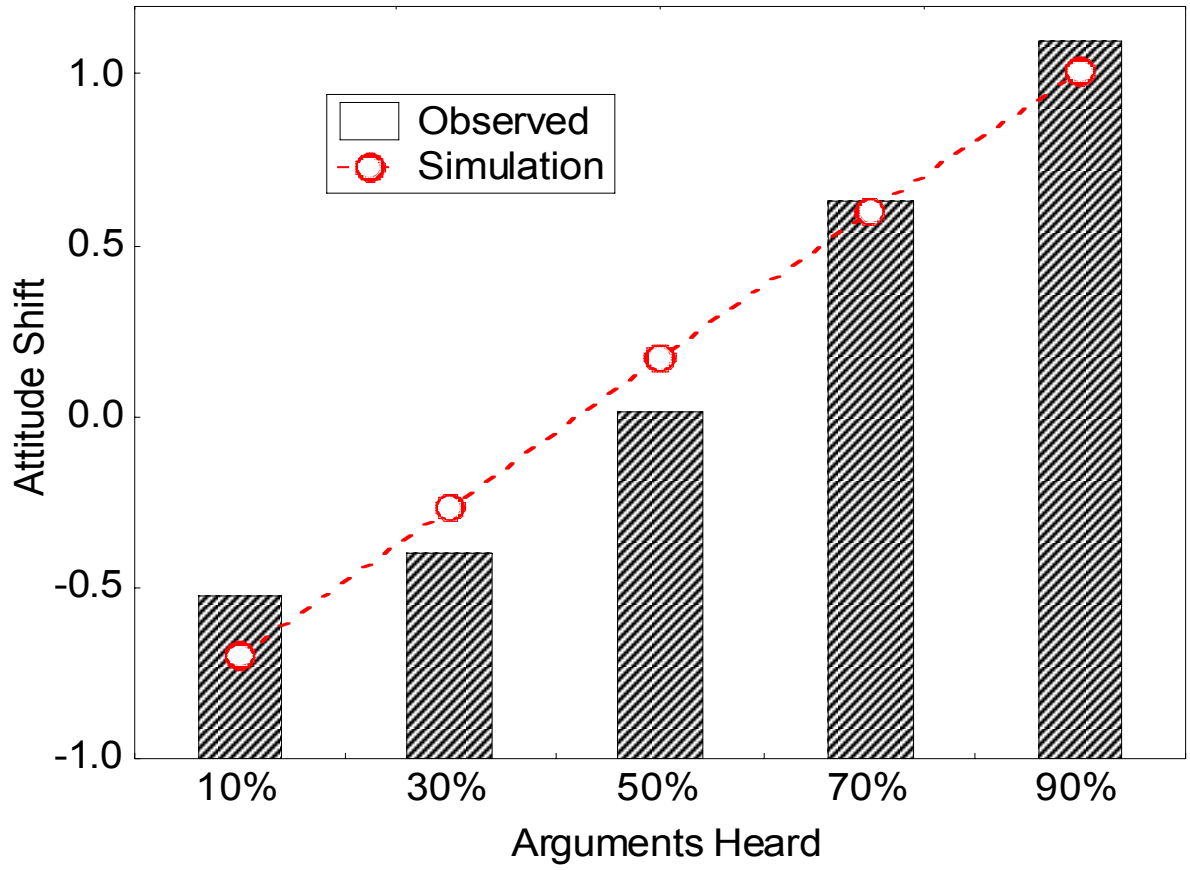

Observed
Simulation
Attitude Shift
1.0
0.5
0.0
-0.5
-1.0
10%
30%
50%
70%
90%
Arguments Heard

Figure 6

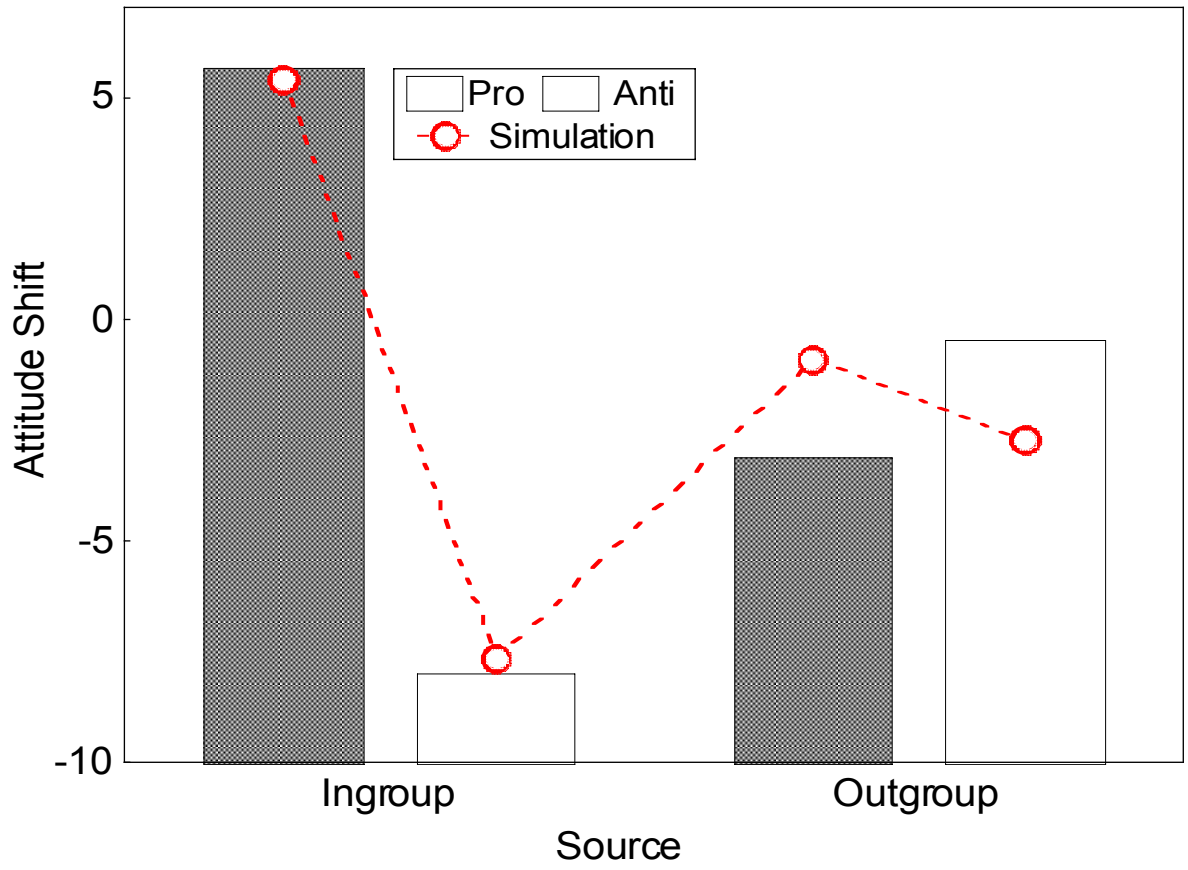
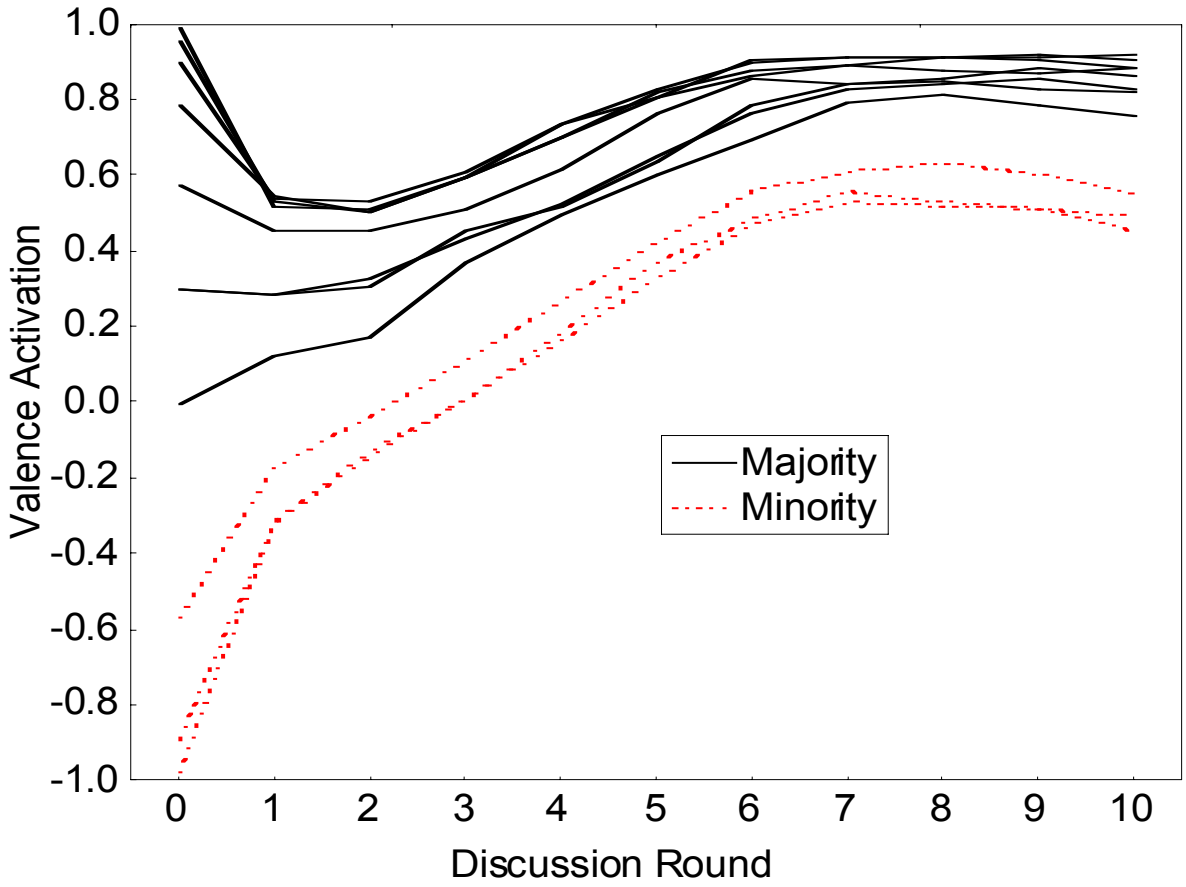
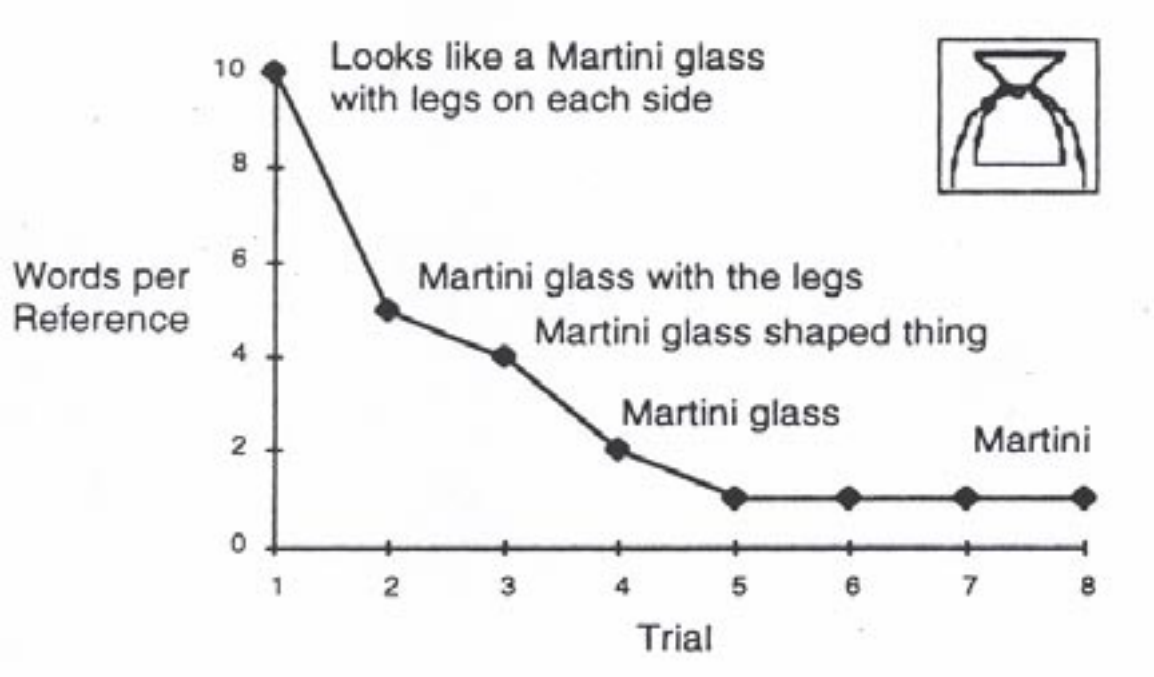

Pro
Anti
Simulation
Attitude Shift
5
0
-5
-10
Ingroup
Outgroup
Source

Figure 7


1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Valence Activation
0
1
2
3
4
5
6
7
8
9
10
Discussion Round
Majority
Minority


Looks like a Martini glass
with legs on each side
Words per
Reference
Martini glass with the legs
Martini glass shaped thing
Martini glass
Martini
10
8
6
4
2
0
1
2
3
4
5
6
7
8
Trial

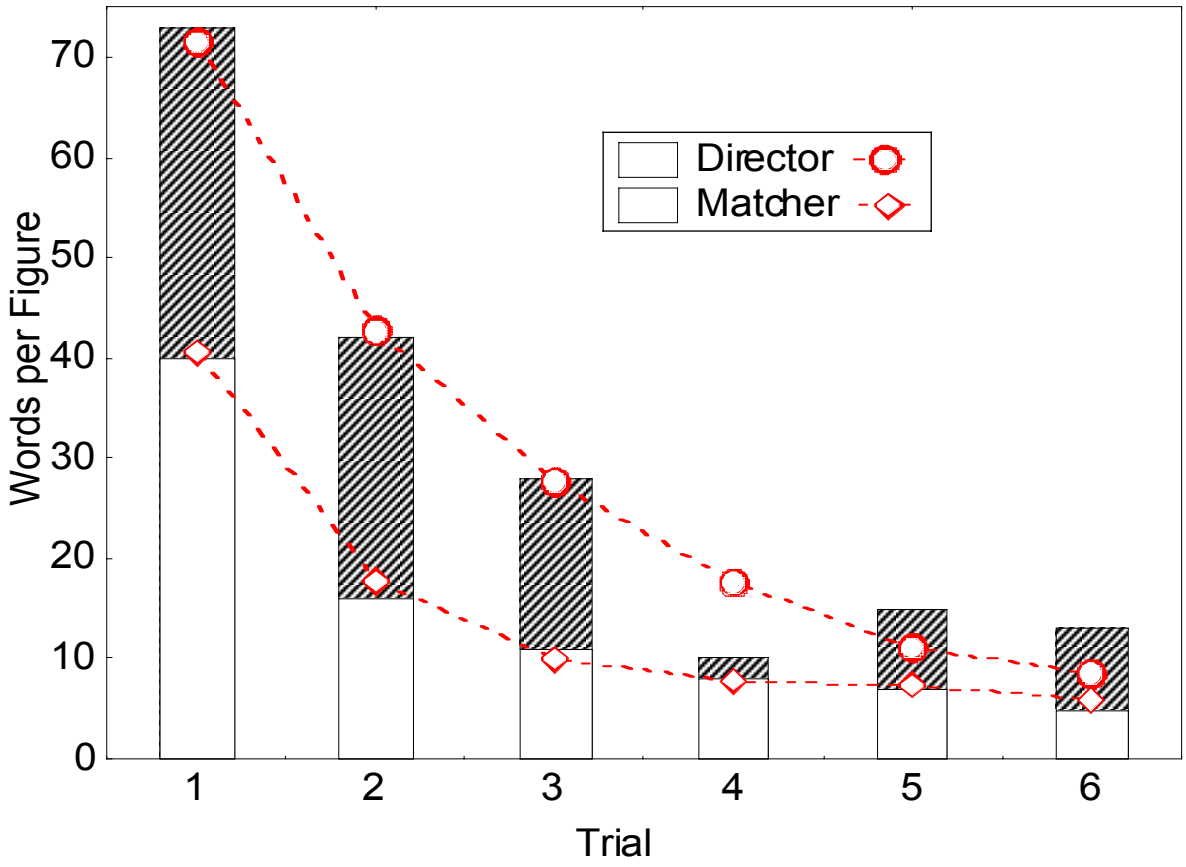
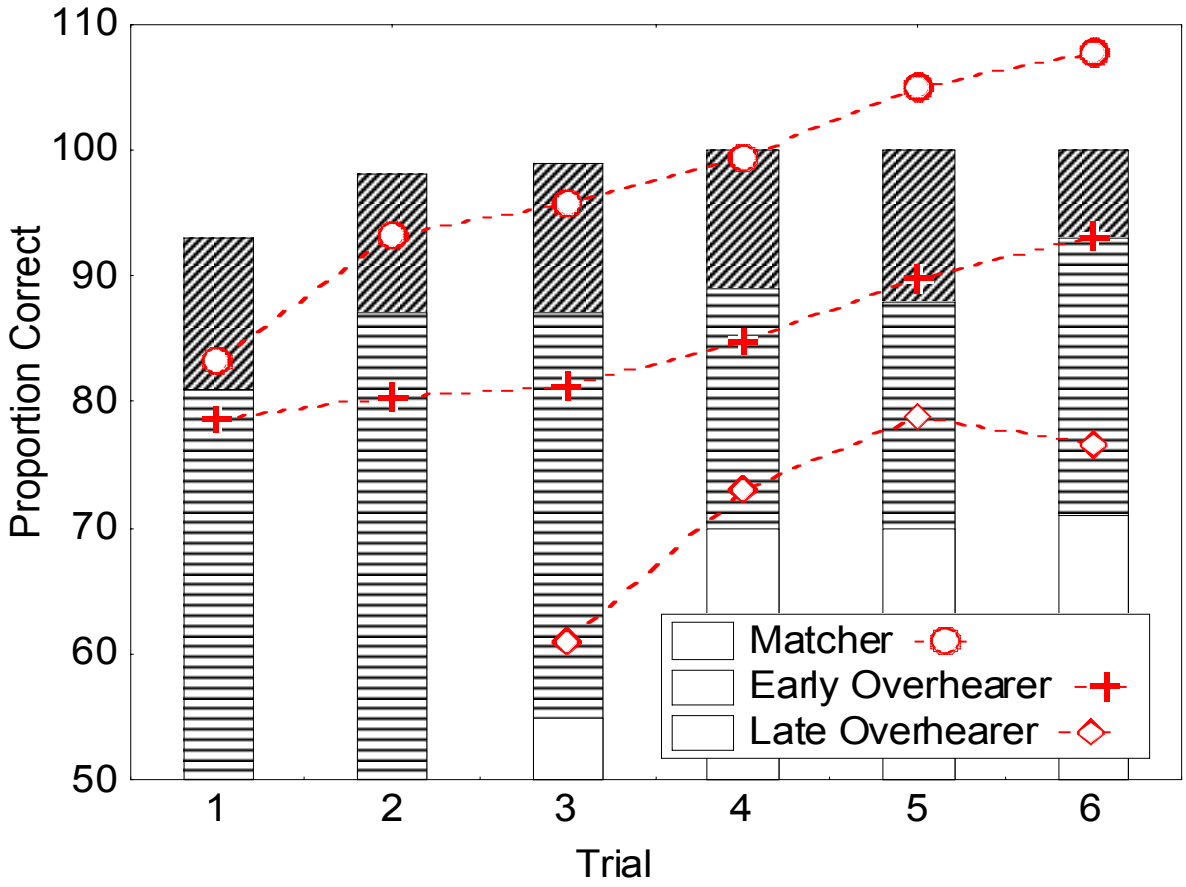
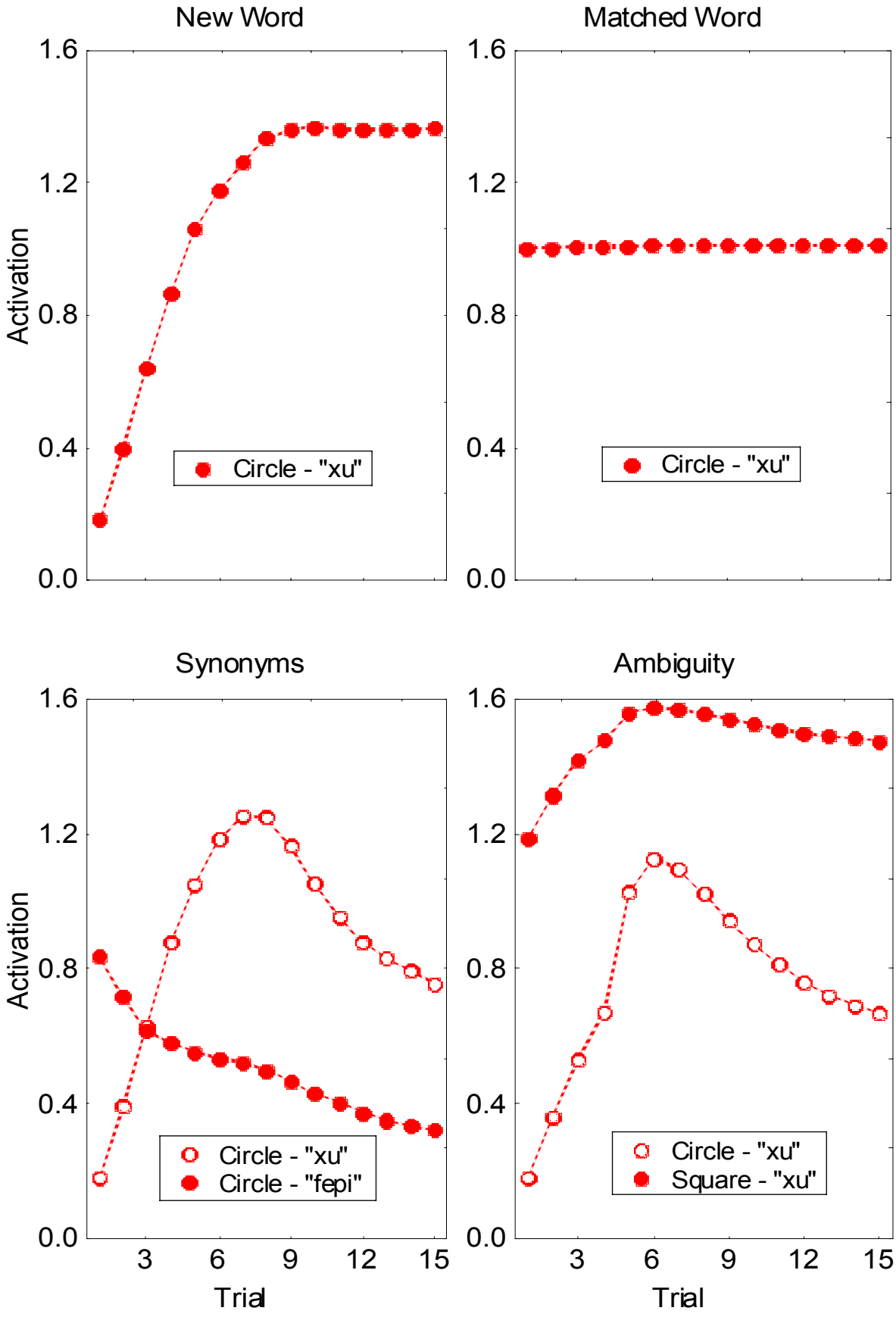

Director
Matcher
Words per Figure
70
60
50
40
30
20
10
0
1
2
3
4
5
6
Trial

Figure 9


110
100
90
80
70
60
50
Proportion Correct
1
2
3
4
5
6
Trial
Matcher
Early Overhearer
Late Overhearer


New Word
Matched Word
Synonyms
Ambiguity
Activation
Trial
1.6
1.2
0.8
0.4
0.0
3
6
9
12
15
Circle - "xu"
Circle - "fepi"
Square - "xu"

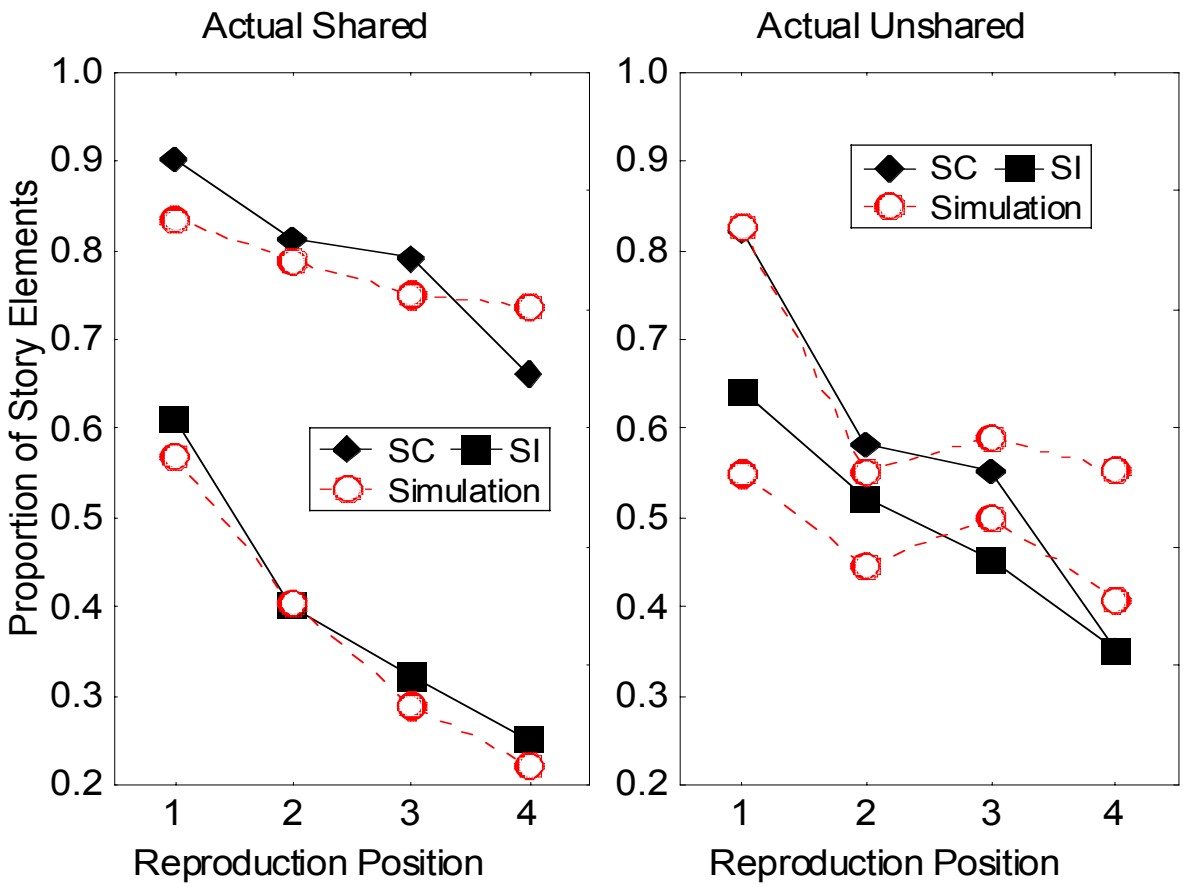
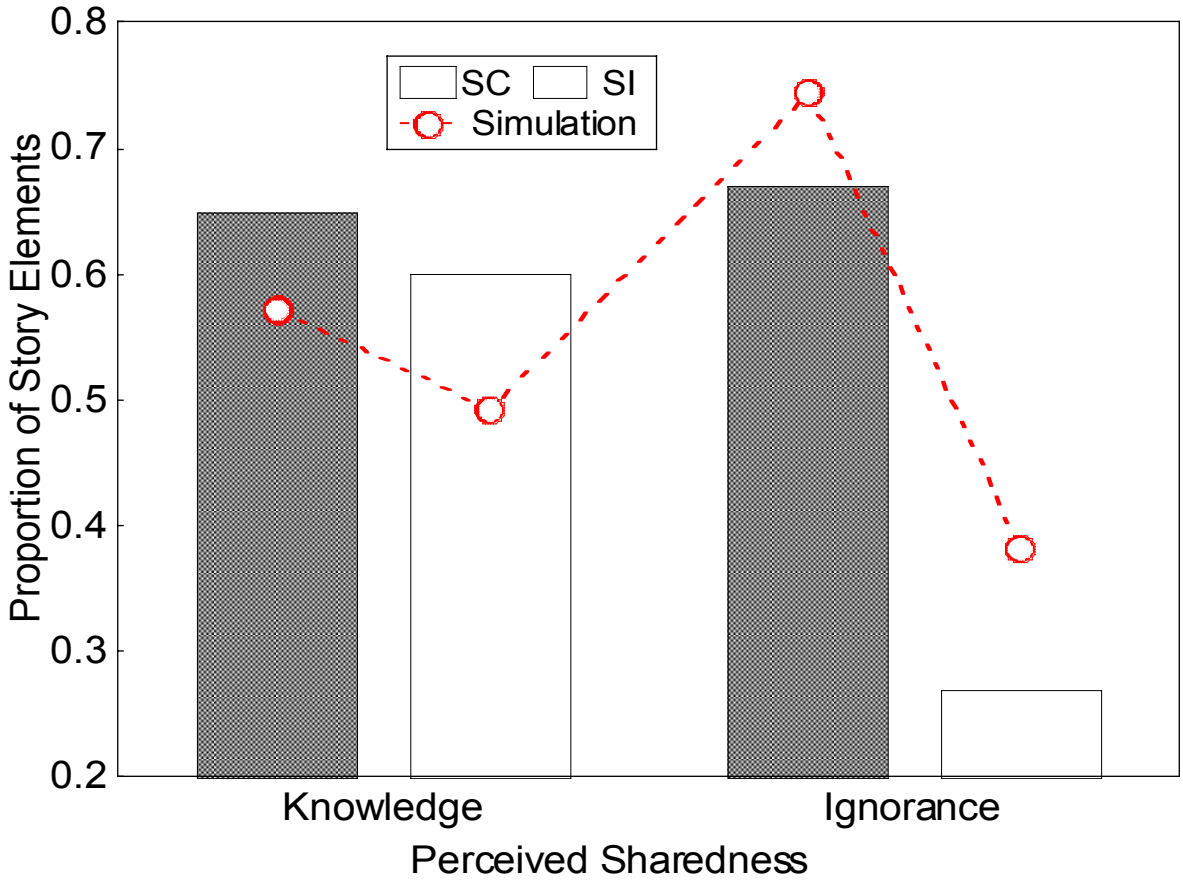
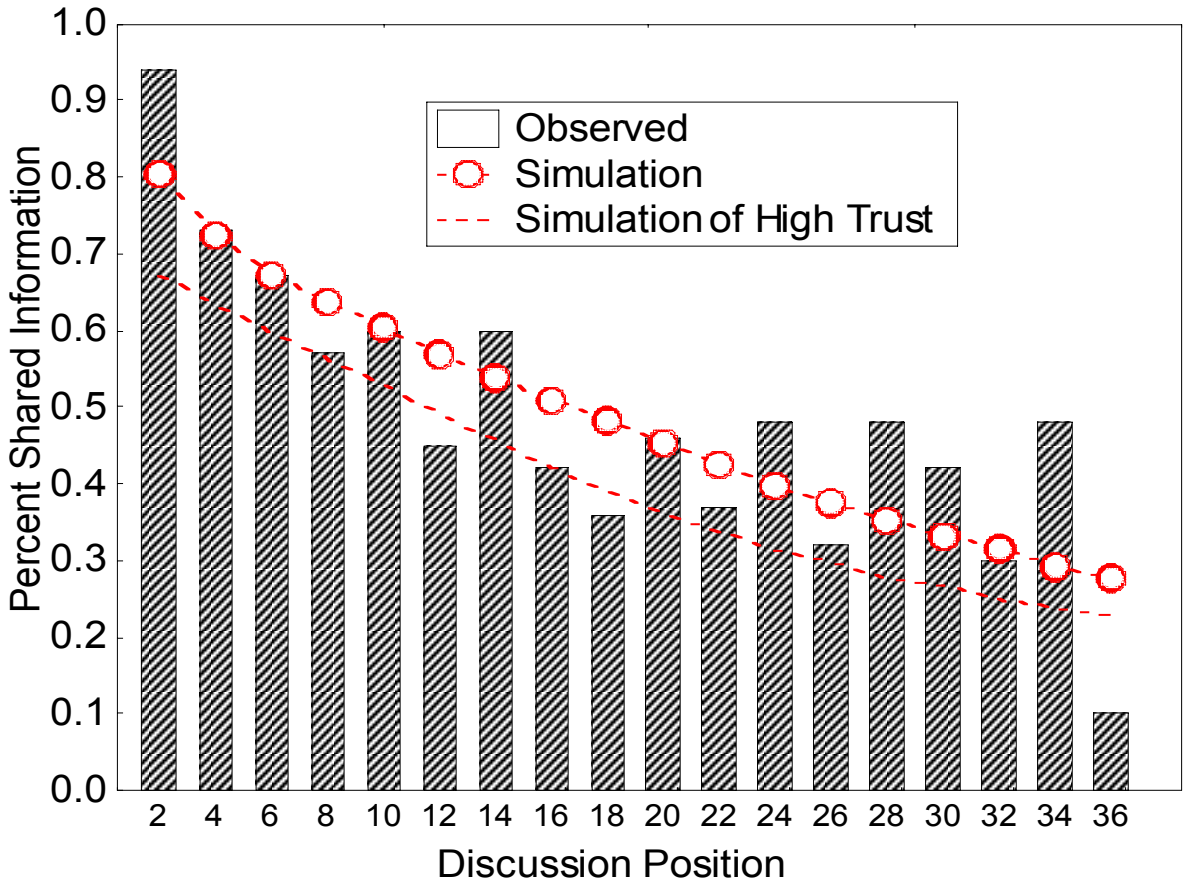
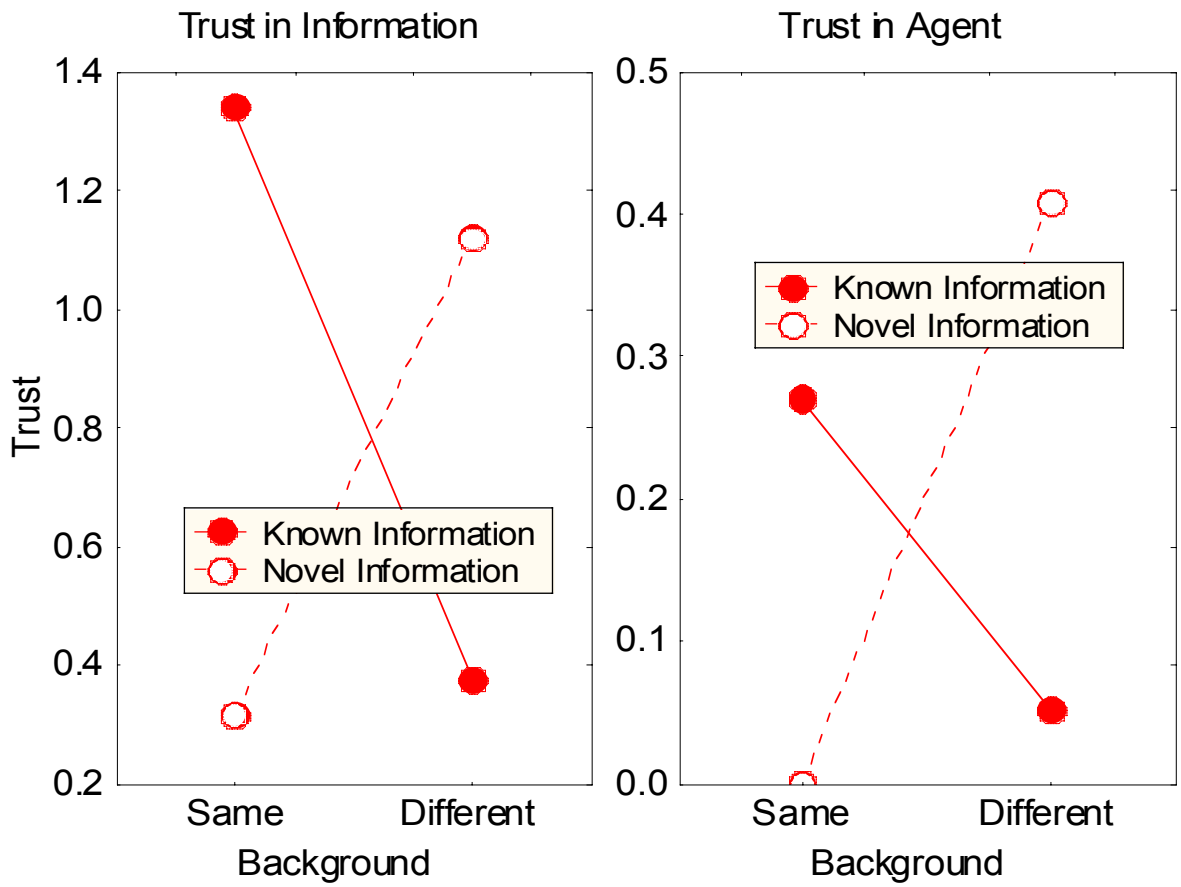

Actual Shared
Actual Unshared
Proportion of Story Elements
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1
2
3
4
Reproduction Position
SC
SI
Simulation

Figure 12


0.8
0.7
0.6
0.5
0.4
0.3
0.2
Proportion of Story Elements
SC
SI
Simulation
Knowledge
Ignorance
Perceived Sharedness


Observed
Simulation
Simulation of High Trust
Percent Shared Information
Discussion Position
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36


Trust in Information
Trust
1.4
1.2
1.0
0.8
0.6
0.4
0.2
Known Information
Novel Information
Same
Different
Background
Trust in Agent
0.5
0.4
0.3
0.2
0.1
0.0
Known Information
Novel Information
Same
Different
Background