Simulating rational social normative trust, predictive trust, and predictive reliance between agents

Maj Tuomela[1] and Solveig Hofmann[2,*]
[1]*Department of Philosophy, University of Helsinki, P.O. Box 9, 00014 University of Helsinki, Finland*
E-mail: maj.tuomela@helsinki.fi
[2]*Department of Philosophy, University of Munich*
*Address for correspondence: Communicate! Gesellschaft für Informationssysteme mbH, Mittenwalderstr. 63, D-81377 München, Germany
E-mail: Hofmannsol@aol.com

Abstract. A program for the simulation of rational social normative trust, predictive 'trust,' and predictive reliance between agents will be introduced. It offers a tool for social scientists or a trust component for multi-agent simulations/multi-agent systems, which need to include trust between agents to guide the decisions about the course of action. It is based on an analysis of rational social normative trust (RSNTR) (revised version of M. Tuomela 2002), which is presented and briefly argued. For collective agents, belief conditions for collective agency should be added. For the various forms of trust agents must have (at least) subjectively rational reasons to believe that the conditions of the trust account are fulfilled. A list of such reasons (of varied weights), e.g., given by empirical research, can manually be built into a parameter file or be generated by a calling program in a fixed format. From this list of reasons the program randomly generates a belief base for the agents of the artificial society. Reasons can be chained together so that one set of reasons satisfies several belief conditions. The program checks if the conditions are fulfilled for the artificial agents' social normative trust/predictive 'trust'/'predictive reliance' in another agent that he will perform an action X. Each outcome is logged to a result file. In conclusion we discuss various aspects of the application of a trust component of the suggested kind in empirical research, social simulation, and multi-agent systems.

Key words: e-commerce, multi-agent systems, reliance, social simulation, trust

Abbreviations: MAS – multi-agent system; RSNTR – Rational social normative trust

Introduction

Simulating trust

Trust is needed in private relations, in groups, between groups, also in the framework of organizations, institutions and states. The development in computer and information technology at the end of the last century allowed for the creation of artificial societies under different aspects of research. Computer simulation was established as a method in social science. The autonomous agent paradigm that originated in artificial intelligence changed against the background of the formation of computer networks in applied computer science from isolated action of single agents to social interaction among intelligent agents of a multi-agent system (MAS) (Castelfranchi and Tan 2000). Through the use of multi-agent models to simulate social systems multi-agent simulation was established as a special application of MAS (Gilbert and Troitzsch 2000).

Recent developments in information technology and electronic commerce have increased the demand to consider the issue of trust online. When using artificial agents to make contracts and agreements, situations arise where the negotiating partners have never met face-to-face. The anonymity of the human agents behind provides new opportunities related to hackers and fraud. The establishment of security systems can help to prevent online communities from attacks of outsiders. However, there is a risk of fraud and deception among the citizens of online communities (Nissenbaum 2001). When a user gets access, there are no sharp criteria for deciding by mere security mechanisms whether he is acting with goodwill or fraudulently. In these situations trust is called for. Here one can take the counterpart of human trust in the case of artificial agents to be a rational decision criterion

and include trust in mediating agents in the decision rules of agents of a multi-agent system. In the area of multi-agent simulation, trust is a crucial factor in many aspects of social simulation. Cases in point are cooperation between single agents, the formation of groups, or the influence of technical development, such as electronic commerce, on societies.

In this paper we introduce a program for the simulation of rational social normative trust, predictive 'trust,' and predictive reliance between agents. We start by giving a brief presentation of different kinds of trust. In the second section the theoretical base of the simulation, M.Tuomela's account of rational social normative trust (RSNTR) is presented and briefly defended. The third section deals with the architecture and implementation of the simulation program. In the last section we discuss questions of its application to social science, to multi-agent simulation, and to multi-agent systems and related ethical aspects. The aim of this project has been to show how a working simulation of trust between agents can be designed based on the presented account of trust.

Trust

The account of rational social normative trust (RSNTR) below (revised version of M. Tuomela 2002) and the analyses of its close cognates purport to satisfy the following criteria. *1. Genuine trust is distinguished from predictive 'trust,' without restricting trust to the area of close relations and non-instrumental motives. 2. Both kinds of trust are distinguished from 'predictive reliance' that allows for accidental gratification. 3. Holding someone as trustworthy is distinguished from trusting. 4. Trusting is distinguished from deciding to depend on someone.* It is claimed that the analysis of (RSNTR) and the analyses of its nearby notions satisfy the above criteria by providing a detailed explication of the following view of trust and its close cognates.

When an agent A has *rational social normative or rational genuine trust* that another agent B will perform an action X, he expects B's intentional good-willed performance of X and believes that he has a social right to expect this due to their relationship of mutual respect, and he expects B's performance of X at least in part because of B's responsiveness to A's right. A's beliefs are based on subjectively rational reasons. On the grounds of these beliefs A feels comfortable about his dependence on B for X. On the grounds of the previous he has an accepting attitude vis-à-vis his dependent position. He may then *decide to depend* on B for X, without considering it risky.

On the other hand, when A has *predictive 'trust'* in B for performing X, he simply has a reason-based

belief of B's intentional good-willed performance of X. On these grounds he feels comfortable being dependent on B, and on the grounds of the previous he has an accepting attitude regarding his dependent position. A can then *decide to depend* on B for X, feeling secure. The difference between these two kinds of trust lies in the social normative dimension of genuine trust, which is due to the relationship of mutual respect between the agents. Relationships involving a socially or (quasi-) morally grounded 'we-perspective' or 'you-perspective' such as friendship, sincere agreement, a relationship governed by mutually respected social norms etc., give us social rights grounded in the relationship to expect good-willed behavior *of* each other, and not only reason for believing *that* the other will behave so. Predictive 'trust' could be said to amount to reliance on another person's qualities or features of the situation, disregarding the trustor's social right embedded in their relationship of mutual respect to have the trustee's responsiveness to general social rights involved in that type of a relationship. When a relationship of mutual respect is missing altogether the trustee, of course, has to do with merely predictive 'trust.' The trustor may depend on features of the situation, such as common norms of decency and on features of the trustee, such as his moral qualities or fear of social sanctions. (Cf. Hollis 1998: 160 for a distinction between normative trust (expecting of) and predictive trust (expecting that); and Loukola 1999: 21–23 for a discussion on this topic.)

A's evaluation of B's trustworthiness differs from A's trusting in B by the fact that the former consists of beliefs only, while the latter also involves A's feeling of comfort at the prospect of being dependent on B and A's accepting attitude regarding his dependent position. Here A does not trust B, but he may decide to depend on B for X due to his evaluation of B's trustworthiness. Then A could be said to *act as if he trusted.*

A's *'predictive reliance'* on B for X is a prediction of B's intentional good-willed behavior toward someone else or nobody at all, vis-à-vis X, or his non-intentional gratification of A or someone else. (Cf. Gambetta 1988, 218 for a definition of trust that allows for the trustee's unintentional gratification of the trustor.) A may be dependent on B for X or is only interested in what B will do. *Deciding to 'trust' is to act as if one trusted.* Trusting is nothing one can do at will. The trust account below and the analyses of its close cognates solve the major problems concerning trust, discussed in relevant literature (See criteria 1.–4. above).

The rationality of the attitude of trust is a subjective rationality, and trusting is claimed to be rational when the trustor has the beliefs specified in the account

(RSNTR), or a disposition to have such beliefs. These beliefs are based on subjectively rational reasons or dispositions to have such. Both the beliefs and the reasons may be unreflective. Attitudes are subjective and people's reasons for trusting are subjective, due to varying resources and levels of optimism. For 'objective' predictions we would need 'objective' criteria for what counts as a good reason for a specific belief. Empirical research could only offer *typical* reasons for trust.

The present trust account may be applied to collectives when the conditions for collective agency are satisfied (see e.g., R. Tuomela 1995, Chap. 5 and R. Tuomela and M. Tuomela 2003; for collectives' trust in collectives see M. Tuomela 2003b). Then A is a collective that can have binding collective beliefs and other attitudes and B is a collective that can have binding collective beliefs and intentions, and ability to act as a collective. This ability involves the collectively committed members' control over 'soloing' members, and in case of failure, a collective responsibility for the action.

The account of trust

Rational social normative trust (RSNTR)

The account of rational social normative trust (RSNTR) for primary trust relations is presented below, starting with the account of the context for rational trust (Y).

The following conditions are necessary and jointly sufficient for a context (Y), 'the trust context,' in which A can rationally trust B that he will perform a specific action X:

a) A wants B to perform X, an action that A does not intend to perform himself. A thus believes that he is concerned with B's performing X, and believes that he is in that sense dependent on B's action;

b) A has a belief about B's belief of the dependence of A; and

c) A believes that his dependence is 'relevant' to B. That the dependence is relevant to B means here that B is free to let A's dependence play a role in his decision-making, and that B takes A's dependence into account in his actions.

A's beliefs in a)–c) may be dispositions to have such beliefs and they are based on (at least) subjectively rational reasons (or dispositions to have such).

(RSNTR) For a person A to have rational social normative trust in another person B that he will perform a specific action X, in a situation in which context Y is present, the following necessary and jointly sufficient conditions must hold:

1. Intention condition: A expects (or has the disposition to expect) that B intends to perform X. The expectation is based on (at least) subjectively rational reasons or dispositions to have such.

2. Ability condition: A expects (or has the disposition to expect) that it is possible for B to perform X (viz., that the internal and external opportunities for B's action obtain). The expectation is based on (at least) subjectively rational reasons or dispositions to have such.

3. Goodwill condition: A expects (or has the disposition to expect) that B will intentionally act with goodwill, including goodwill towards A, when performing X. A believes (or has the disposition to believe) that due to their relationship of mutual respect he is entitled to expect this of B, on social normative or (quasi-) moral grounds, and that B acts at least in part because A has this right. These expectations (or dispositions) are based on (at least) subjectively rational reasons or dispositions to have such. The reasons for A's expectation in 1 may be (partly) the same as here.

4. On the grounds of 1, 2 and 3, A has a 'positive' feeling. This feeling is or, at least, includes the feeling of being comfortable about his dependence on B for performing X.

5. On the grounds of 1, 2, 3, and 4, A has an accepting attitude vis-à-vis his position of dependence on B for performing X. This attitude is non-intentionally acquired.

The rationality of *social normative or genuine trust* is defined by (RSNTR) 1-5 and the trust context *Y.* B's 'acting with goodwill' involves, firstly, his acting with special commitment to his performing X. Secondly, it involves his acting intentionally so as to advance A's interests, with a description of interests that at least partially coincides with A's description. This is considered the same as B's acting with a pro-social attitude, or his being at least one-sidedly socially committed to A to perform the action. Acting with goodwill accounts for the persistence of B's intention and for his acting with persistence and flexibility in a way that at least does not harm A.

It should be noted that the belief of B's 'acting with goodwill' does not mean that B is necessarily believed to be genuinely good-willed towards A. To intentionally act with goodwill may be based on genuine caring or respect, but also on prudential or moral reasons. In rational social normative trust, the trustor has rational reasons due to their relationship of mutual respect to normatively expect of the trustee, on social or (quasi-) moral grounds, that he has a 'we-perspective' or a 'you-perspective,' or more specifically, that A has rational reasons to believe that B has such a relationship to A that when B performs the action, he ought to

take into account 'their' point of view or A's point of view. As this is how A sees the relationship, he normatively expects of B, on social or quasi-moral grounds, that B will intentionally act with goodwill towards A. A also expects that B will do so at least in part because A has this right.

A can be wrong about what their relationship is, about the norms of decency connected to various relations, and about B's awareness of the norms. If A believes that B is not willing to act in the right way, A may think that he is mistaken about their relationship, e.g., a friendship. This is how the 'I thought you were my friend' – situations may start. If there is an agreement between A and B that B will perform X, and A believes that B will not do so, A will probably not believe that he is mistaken about the agreement. Instead, he might believe that there was no sincere agreement. A may think that B is a person with whom A cannot have a sincere agreement-relation. A is entitled to expect the performance of X of B, but there are no rational reasons that are grounded a relationship of mutual respect for that expectation. When A believes that his relationship with B is based on mutual respect for prudential reasons, the respect may be genuine, but A may believe that B's goodwill in A's expectation of B's good-willed action is restricted to respecting A's rights. In this case A's genuine trust is 'thin.'

Cases of rational 'predictive reliance' or prediction which do not qualify as rational social normative trust, or any other kind of trust, are analyzed as in (RSNTR) with the following alternations: One or several of the beliefs in Y are lacking or Y concerns B and a third person C, A having the beliefs; (RSNTR) 4 and 5 are lacking; the goodwill-expectation concerns C or nobody at all, and no socially grounded normative expectations are required. Rational predictive reliance is comparable to prediction of another person's action. The action may be accidentally beneficial to the relier, who may be dependent on the action. When it is intended to be beneficial to a third person or to the actor himself, and the relier is unconcerned with the action, he is an outsider and has only sufficient interest to evaluate the situation.

The trust account (RSNTR), above concerns rational social normative trust, which is regarded as genuine trust, as opposed to rational predictive 'trust.' The latter kind of 'trust' is analyzed as in (RSNTR) with the exception of the socially or quasi-morally grounded normative expectation. In rational predictive 'trust' A only expects that B will perform X, intentionally gratifying A, while still having the beliefs indicated in (RSNTR) 1–3 (without the entitlement belief), whereas in rational social normative trust A also expects of B that he will do so, and indeed, that

he will do so, in part for the reason that he has a social right to B's gratifying action.

Thus, in rational predictive 'trust' A predicts that B will intentionally gratify him by performing X, feels comfortable about being dependent on B and has an accepting attitude concerning his dependent position, e.g., due to B's moral character, fellow-love, or his fear of legal sanctions. In predictive 'trust' A merely relies on some features of the person or the situation, instead of also on reasons pertaining to his social rights involved in their relationship of mutual respect. Genuine trust requires more, which is expressed in (RSNTR) 3 by A's socially or (quasi-) morally grounded normative expectation, due to the relationship of mutual respect between A and B, and by A's belief that his right is a partial reason for B's action.

Holding someone as trustworthy vis-à-vis the performance of a specific action, is to have the beliefs in (RSNTR) 1–3, with, or respectively without, the normative expectation. For trusting we need to add (RSNTR) 4–5, in order to include the affective component and the accepting attitude concerning a dependent position. Both in predictive and social normative trust we may decide to depend on another for an action, or refrain from doing so. When A decides to depend on B for X, having merely evaluated his trustworthiness, A acts *as if* he trusted B. When A has genuine (social normative) or predictive trust in B, and decides to refrain from depending on B for X, he actually trusts B, although he chooses to e.g., perform the action himself. In predictive 'trust,' however, the trust is not much more than counting on another person's qualities and motives or some situational features.

A brief defense of RSNTR

The analysis draws a line between trust and a broader notion of reliance through the *trust context (Y)*. Y does its work with a view of trust that involves the trustor's (A) expectation of the trustee's (B) *intentional gratification* of A. When one or several conditions for Y are lacking between A and B, A can only consider whether or not to have 'predictive reliance' that B will perform X. If A is dependent on B's action, he may expect B's unintentional gratification of himself as a side effect. In our opinion these cases do not deserve the label of trust, but fit the more general notion of reliance. Trust is an attitude that presupposes that there is a relationship between the parties. To expect that someone's action will be beneficial to oneself by accident is just part of predicting the consequences of his action. The use of 'reliance' as a general term is defended by its wide usage, from expressing one's views of future

events and the safety or functional effectiveness of objects to trusting a person. Although reliance is sometimes used as a synonym for trust, it is here used as a technical term for a prediction-like evaluation. The relier is either dependent on the action of the person relied upon and can only expect to be gratified as a side effect of that person's action or he simply takes an outsider's position evaluating the situation in which a third party may be involved.

To Gambetta trust belongs to a context of uncertainty (Gambetta 1988: 218). We agree with him that there is uncertainty, but only in the sense of Y c). The trustee is free to and will take into account in his action the dependence of the trustor in whatever way he chooses to. Thus, we agree with Lagerspetz that one does not consider it taking a risk when one trusts (Lagerspetz 1998: 2). Trust involves believing that one is in good hands. So, with the reservation of the uncertainty due to Y c), we agree with Pettit that trust may figure in contexts of certainty (Pettit 1995, 204).

The necessity of A's belief of his dependence on B's action (Y a)) is based on an intuition about the meaning of 'trust' that is shared by many scholars. An 'unconcerned' relier is merely predicting the course of action of the other person. In Y a) A's 'dependence' is defined as meaning that A is concerned with B's action in the sense that "A wants B to perform X." It is claimed that it would be irrational for A to trust in B that he will perform X when A believes that he is unaffected by B's performance. Trusting is an investment, and there should be a cost to A, if his trust is betrayed. There is no cost, if A is unconcerned and mistaken in his reliance. If B's action is unwanted by A, he is not in a context to consider trust. E.g., "I trust that you will do what I fear you to do," does not fit the idea of trust. We do not need a notion of dependence that is stronger than A's 'want' as the trust account involves *necessary and jointly sufficient conditions*. To require that A has a goal that B performs X is more than is needed here, and also brings us to a controversial topic. Even aim-goals require some disposition towards action on the part of the goal-holder. A trustor need not have that disposition, and he should not have it as he believes that he will not be betrayed when he trusts. (Cf. Falcone and Castelfranchi 1998 on a trust account stressing B's goal adoption, where goals include wants.)

Y b) is needed, as B might unintentionally harm A, although he is good-willed towards A, if B lacks a belief of A's dependence. Y c) is necessary for the reason that B has to be free to make A's dependence a reason for his action. If he is bound to perform X anyway, A is in a predictor's position. Also, A has to believe that his dependence is relevant to B for expecting B's intentional gratification of A. If A believes that his dependence does not figure in B's

plans, A cannot expect B's intentional gratification of A.

(RSNTR 1), the *Intention condition*, is needed for rational trust, as A needs to expect that B intends to perform X to be able rationally to trust that B will perform X. (RSNTR 2), the *Ability condition* is needed as A also has to expect that it is possible for B to perform X, in order to rationally trust that he will perform X.

(RSNTR 3) The *Goodwill condition*. Without expecting that B will intentionally act with goodwill, including goodwill towards A, when performing X, A cannot trust that B will perform X at all, although he could predict that B is likely to perform X. Firstly, if B's commitment to his performance of X did not include a commitment to A when performing X, he could more easily stop intending to perform X when circumstances change. The social commitment is crucial, as it binds B to the performance of the action with a 'glue' made of strong ingredients. In the case of genuine trust (RSNTR), A also believes that A has a right due to their *relationship of mutual respect* involving this right, to expect B's good-willed performance of X, or B's social commitment to performing X, whether A has expressed his expectation or not, which will be argued below. A also believes that B will act at least in part because A has this right. The relationship of mutual respect may e.g., be based on caring, social and moral decency, or just be based on sincere agreement, 'The one-sided social commitment' means that A's acceptance of B's commitment is not required, as in accepting a promise. A simply believes that B ought to be and is socially committed to A in part because A has this right due to their relationship of mutual respect.

Secondly, without goodwill-expectations A cannot believe that B will perform X in a way that avoids harming A. When we say that we trust someone we intuitively mean that he will act without harming us. Thirdly, A may think that B could 'arrange' circumstances in a way that excuses him from acting, if A has reasons to believe that B will act without goodwill towards A. A's expectation of B's acting intentionally without ill will is the borderline case for A's expectation of B's intentional good-willed action. The *'restricted freedom of the trustee'* is required for rational trust. We do not trust people, if we expect that their view of our good differs from our own view, although they are free to advance our interest as they see it. A goodwill-aspect or some notion of 'unharmfulness' is included in many accounts of trust. (Cf. Baier 1994: 98–99.)

(RSNTR 4) acknowledges the positive feeling that goes with trust. It may vary from warm feelings towards the trustee to feeling comfortable about one's

dependence on him for performing the action. It is argued that it would be non-rational not to feel good when one has reasons to feel so. (RSNTR 5) establishes that A is in the trustor's position through his having an accepting attitude regarding his position of dependence on B for performance of X. This distinguishes trusting from *believing that someone is trustworthy* and feeling good about it.

However, what ultimately distinguishes a person's trusting someone from his believing certain features of him and his predicting the person's behavior along with having a comfortable feeling about his position of dependence and his accepting attitude of such a position, is that in genuinely trusting the person he expects the good-willed behavior *of* him because of their relationship of mutual respect, and that he will act in part for the reason of the trustor's socially grounded right embedded in their relationship, and not only *that* he will act with goodwill merely because of some other reason, e.g., friendliness or general norms pertaining to that type of situation.

The key lies in the relationship of mutual respect. In trust, the relationship has to be such that the trustor believes that he has a socially justified right to expect good-willed behavior of the trustee. He must have rational reasons to believe that their relationship is such that the trustee ought to have, on social or (quasi-) moral grounds, and indeed has a you-perspective or a we-perspective. E.g., the trustee is or aspires to be a friend, a member of the same group, a respectful party of a joint venture or a simple exchange. In these examples the relationship is such that the trustor has a social right to expect the trustee's good-willed behavior of various degrees in varying situations. Of course, the trustor also needs reasons to believe that the trustee understands their relationship in the same way, has a similar view of social norms pertaining to specific relationships, and acts as he should at least in part because the trustor has a right to expect the trustee's good-willed action. It should be stressed that the social right is due to the relationship of mutual respect and not only to some common norms of decency that apply to the situation. The social-normative expectation and the belief of the trustee's acting for the right reason are needed for rational genuine trust, and these are included in (RSNTR 3).

On the other hand, when a person predicts that another person will save his life, e.g., due to his moral character, or fellow-love, he is simply predicting future events and feels good about being in the hands of the person, having an accepting attitude about being vulnerable. This we have called *predictive 'trust,'* as opposed to social normative trust. In predictive 'trust' the trustor merely depends on some features of the trustee or the situation. The opposition against

predictive 'trust' as deserving to be considered trust, at all, is widespread in the literature.

The discussion about predictive and normative trust constitutes the defense of the necessity of the normative expectation in genuine trust. Thus, the necessity of the conditions of the trust account has been briefly argued for. The joint sufficiency of the necessary conditions is suggested for the following reasons. There seems to be wide agreement about the relevance of such components of trust as beliefs of intention, ability, persistence, gratification, and affect. What has here been argued to be missing has been added 1) through context Y and the intentionality requirement in the Goodwill condition. This draws a line between trust and a prediction-like reliance, viz. between trust and an evaluation of what others will do that affect others or has some accidental effects on the relier; 2) through the fifth condition. This shows the distinction between holding someone as trustworthy (in a normative or predictive sense) and having the attitude of trusting; 3) through the trustor's social-normative expectation in the Goodwill condition and his belief of the trustee's acting in part because of the trustor's right. This normative expectation grounded in a relationship of mutual respect supplies the final distinction between genuine trust and predictive 'trust.'

'*Deciding to trust*' sounds like making a bet on someone, or *acting as if one trusted.* When we trust we think we will not be betrayed. Also, trusting is non-intentional. We either have the attitude or we do not. However, when we trust, in a normative or predictive sense, we may decide to depend on the trustee or refrain from doing so. If we choose to depend on him, this is based on genuine or predictive trust. In the latter case, the 'trust' is comparable to reliance on some personal or situational features of the trustee. This hardly deserves to be called trust, at all.

Applying the trust account in e-commerce

Rational social normative trust (RSNTR) requires a relationship of mutual respect between the trading parties in the context of e-commerce. As the seller gets the credit card number before he delivers the goods he is less vulnerable than the buyer. What reasons then could the buyer have to believe that the seller will respect his rights? A well-known company can both be traced and is careful with its reputation. However, when a person buys something from an unknown company he may not have strong rational reasons for believing that the company will treat him with respect. In predictive 'trust' the trustor depends on features of the partner and the situation. He may find it likely that a well-known company will deliver the goods, due to

fear of sanctions: the company is traceable and it will endanger its reputation if it cheats. If every company selling something through the net could be traced down, cheating could be legally and socially sanctioned. A black list of untraceable cheating companies could have a deterring effect on cheating as well as be a direct help to online buyers. Helen Nissenbaum discusses the issue of security and trust in cyberspace and points out that high security for parties in online communication leaves no place for trust and may even undermine a climate of trust (Nissenbaum 2001). We fully agree that trust presupposes the freedom of the trustee to disappoint the trustor (RSNTR Yc). We also agree that when control measures are effectively introduced a climate of trust will not be cultured. Nissenbaum still grants that we need pockets of high security for e.g., commerce and public institutions. In our terms, security is needed when trust is hard to establish due to the fact that it is difficult to make a rational prediction of an agent's compliance or to create a relationship of mutual respect among the partners in e-commerce. However, as was pointed out above, when reputation becomes valuable, as is the case with any serious company, rational predictive 'trust' and even 'thin' rational social normative trust get a chance. Nissenbaum wants to preserve some more freedom for agents online outside the 'pockets of high security.' This recommendation seems sound. Those who want to communicate anonymously know that the partners may be insincere. However, they may enjoy the gamble, hope for sincerity, and maybe even come to trust. Even letters between strangers may contain clues about goodwill and respect.

Simulating trust between agents

In the following we introduce a program to simulate trust between agents. The basic assumptions of the presented account of RSNTR are translated into program rules of a multi-agent simulation. The activation of agents and the control of the action loop is based on the multi-agent system SMASS, a frame for social simulations (Balzer 1999). In game theoretic approaches trust between agents often is simulated as a special case of the prisoners' dilemma, i.e., a two players' game (e.g., Bacharach and Gambetta 2000). The architecture of a MAS allows for a participation of a variable number of agents, depending on the situation to be simulated. The programming language is Prolog.[1] In the following program description we use

---

[1] The simulation is written in LPA *Win-Prolog* 4.300, Logic Programming Associates Ltd. It runs on Windows 9x, 2000, NT and XP platforms. An executable program with short

a Prolog-like notation: variables start with a capital letter, instances of the variables start with a small letter. Names of Prolog predicates, facts and rules, start with a small letter, too. Symbolic names of variables and mathematical sets are denoted by cursive capital letters.

Program architecture

For rational social normative trust agents must have (at least) subjectively rational reasons to believe that the conditions of the trust account are fulfilled. A list of such reasons, basically to be given by empirical research, can manually be defined or be generated by a simulation program in a fixed format. Each reason can be given a different weight. The predefined settings are passed to the program by a parameter file. From the list of reasons the program randomly generates a belief base for the agents of the artificial society. A set of belief sentences represents the belief of agent a (*trustor*) in a situation where he would need to trust in another agent b (*trustee*) that he will perform a specific action x. The belief base will be generated for the defined set of agents A. It will be kept in memory, but logged to a file as well. The program checks for which agents the conditions for rational social normative trust/predictive 'trust'/'predictive reliance' are fulfilled. The outcome is written to the result file which also contains a protocol how trust respectively predictive reliance was achieved. It can be used as input for a multi-agent-based simulation or multi-agent system if one wants to include trust between agents in its decisions about the course of actions.

Building the parameter file

The set of agents A and the set of reasons have to be specified by Prolog facts in the parameter file. Agents are identified by integer values, so only the number of agents has to be defined by the predicate actors (X). By defining a predicate trustee (Y) a trustee can be chosen among the agents A. The fact actors(8) defines that 8 actors will be generated, trustee(2) indicates that agent 2 will be chosen as trustee. A reason is represented by a structure of the following form: reason(*ReasonId,Trustcondition,Weight,Chain Trustcondition,ChainReasonId,NormIndicator*).

ReasonId assigns a unique identifier to each reason. In the example parameter file these identifiers are built from the letter r followed by a unique number. An interpretation of the identifier will be given if a predicate explain(*Identifier,Meaning*) is specified. For

---

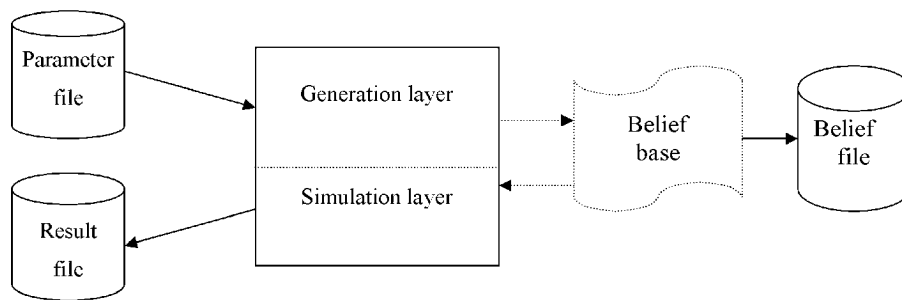description and demo parameter file is available on request at Hofmannsol@aol.com.

Figure 1. Program architecture.

instance the fact explain(r24,b_has_skill_to_perf_x) interprets r24 as 'b_has_skill_to_perform_x.'

The second variable *Trustcondition* specifies in which trust condition *ReasonId* is a reason to create a belief. Valid instances are rsntr1, rsntr2, rsntr3, ya, yb, and yc. Each reason can be given a certain weight to produce a trust belief. The third variable *Weight* will be used for this, e.g., reason(r24,rsntr2,96,0,0,0) will produce a trust belief in RSNTR2 with probability 0.96. By assigning the same *ReasonId* to more than one trust condition with same/different weights the reason will produce belief sentences in this set of conditions with the specified probability.

The variables *ChainTrustcondition* and *ChainReasonId* are used to chain different reasons to produce belief sentences pertaining to different trust conditions at the same time, as will be explained below. If, like in the given example, the value 0 is assigned to both variables no chain will be induced.

The last variable, *NormIndicator*, is used to distinguish between normative and non-normative cases. As argued above the trustor has to have rational reasons to believe that their relationship is such that the trustee ought to have, on social or quasi-moral grounds, and that he indeed has, a you-perspective or we-perspective. Thus a subset of reasons pertaining to RSNTR3 can be marked to produce additional belief sentences related by setting *NormIndicator* to the value 'norm.' The value 'nn' has to be used for the subset of reasons that cover non-normative cases. As this affects only RSNTR3 the variable *NormIndicator* is not used for reasons related to other trust conditions.

The set of agents and their initial trust belief sentences will be randomly generated depending on these settings in the parameter file.

Chaining reasons

Mostly the reasons for the beliefs which lead to a person's trust in another person performing an action X are not independent from each other. They may be based on a certain kind of relationship between trustor and trustee, e.g., friendship or kinship, the

type of cooperation that they are involved in, e.g., agreement-based cooperation or free group activity, or on the trustor's assumptions about the trustee's personality or goals like moral character or concern for his reputation. Thus, one basic reason may pertain to the forming of several of the beliefs in the trust account. In the simulation the reasons for several beliefs are then chained together so that a set of reasons (involving the same basic reason) produces belief sentences pertaining to one or more conditions at the same time. As mentioned above the fourth and fifth variable of the reason-structure are used to chain reasons. *ChainTrustcondition* indicates the chained condition, *ChainReasonId* contains the instance of the chained reason, e.g.

reason(r17,rsntr1,95,rsntr3,r37,_),

reason(r37,rsntr3,95,yc,r67,nn),

reason(r67,yc,95,0,0,_).

Figure 2. Chaining reasons.

The chain of reason r17, r37, and r67 shown above will produce a belief sentence pertaining to RSNTR1, RSNTR3, and Yc at the same time, each with probability of 0.95. Giving the explanation 'b is moral and b believes he should perform x for moral reasons' to r17, 'b is moral and b believes he should perform x for moral reasons and b has no reason to act with ill will towards a' to r37 and 'b is free to perform x and a's want matter to b due to moral reasons' to r67 we have a subset of the reasons to trust in the case of morality. Adding the facts

reason(r27,rsntr2,96,0,0,_),

reason(r43,ya,80,yb,r53,_),

reason(r53,yb,80,0,0,_),

Figure 3. Chaining reasons.

and interpreting r27 as 'b has all resources for performing x,' r43 as 'a wants x and cannot perform x but b can perform x,' and r53 as 'b believes a wants x and cannot perform x, but b believes a believes b can perform x' we have the set of reasons to give the trust conditions in the case of morality. Note that a second chain was introduced between r43 for trust context Ya and r53 for trust context Yb.

The example parameter file consists of a set of (partly) chained reasons involved in twenty-one different cases where trust is considered. The six cases agreement, kinship, same group and trustee has expressed intention, friendship, law, and group norm are examples which can lead to rational social normative trust. The nine cases of liking, reciprocity, reputation, group pressure, social norms, morality, group activity, role-task, and goodness default-value are examples that can lead to predictive trust with more or less probability. The cases of habit, behavior, interest but reason to harm, and predictive reliance involve reasons that can lead to predictive reliance. The cases immoral action and disagreement are examples of distrust which neither lead to trust nor to predictive reliance. This list of chained reasons and their assigned weights given in the example parameter file is based on 'good guesses.' It is intended as frame that is easy to customize. It should be adjusted to the object of study or to the simulated situation. New cases can be added on the basis of the given examples. Cases that are irrelevant should be deleted. The weights that are given to the reasons in the existing chains should be verified by empirical research, e.g., by questionnaires.

Generation of the belief base

From the initial settings in the parameter file the program generates the agents' belief base. A belief sentence is represented as a Prolog predicate of the form bel(*Agent*, *BeliefContent*). *Agent* contains the number of the agent. In our case *BeliefContent* consists of a Prolog fact representing one of the trust conditions that may be a belief sentence as well. E.g., bel(1, bel(2,dependent(1,2,x))) means that agent 1 believes that agent 2 believes that agent 1 is dependent on agent 2 performing action x, which is the representation of trust context Yb for agent 1 concerning agent 2. The generated belief set is consistent according to common epistemic logic. It doesn't contain contradictory beliefs. If an agent does not believe that the relevant trust condition obtains the corresponding belief sentence will not be represented in his belief set.

If a trustee B is defined in the parameter file he will be chosen, otherwise he will be randomly selected from the set of agents *A*. For all agents A, A ≠ B, the following steps will be performed:

1. A reason is randomly selected from the set of reasons pertaining to RSNTR1. The weight of the reason is compared to a random number Rn ($0 < Rn <= 100$). A new value is assigned to Rn by the random generator after each selection. If Rn is less than or equal to *Weight* a predicate bel(A,intends(B,X)) representing the *intention condition* will be asserted. For A = 1 and B = 2 bel(1,intends(2,x)) means that agent 1 believes that agent 2 intends to perform action x. If Rn is greater than *Weight* no belief sentence will be generated.

2. If the selected reason induces a chain of reasons, the generation of chained reasons will be performed first. The chained reasons will be marked as already processed. The other reasons pertaining to the rest of the trust conditions will then be processed, possibly inducing a further chain. The decision process concerning whether a belief sentence pertaining to the trust condition RSNTR2, RSNTR3 and the trust context Ya, Yb and Yc is done in the way described under 1. If in all cases the calculated random number Rn is less than or equal to *Weight* the following predicates are asserted:

bel(A,intends(B,X)).	Intention condition
bel(A,can_perform(B,X)).	Ability condition
bel(A,goodwill(B,A,X)).	Goodwill condition
bel(A,dependent(A,B,X)).	Trust context Ya
bel(A,bel(B,dependent(A,B,X))).	Trust context Yb
bel(A,dependence_is_ relevant(B,A,X)).	Trust context Yc

3. If the reason to produce the belief in RSNTR3, the goodwill condition, was marked as normative, two additional belief sentences of the following form will be asserted:

bel(A,has_right_to_expect_goodwill(A,B,X)).

bel(A,acts_in_part_in_response_to_having_the_ right(B,A,X)).

Every agent is randomly assigned a reason to produce a belief sentence in each trust condition. Whether the belief in the particular trust condition actually is produced depends on the weight of the particular reason. Giving a high weight to a reason yields a high probability that a belief sentence in the associated trust condition is produced. Accordingly, reasons for lack of trust can be specified by assigning a low weight or 0.

Simulating social normative trust/predictive
'trust'/'predictive reliance'

After the initial belief base of the agents has been generated, the program checks whether the conditions for social normative trust/predictive 'trust'/'predictive reliance' are fulfilled. For all agents A, A $\neq$ B, the generated belief sentences concerning trustee B performing an action X will be evaluated.

1. If the belief sentences pertaining to the trust context Ya), Yb), Yc), RSNTR1, RSNTR2 and RSNTR3 for agent *a* are generated, the program deduces RSNTR4 and RSNTR5. We do not want to argue that our artificial agents *have* a positive feeling by doing this. But for simulation of trust it is important that a trusting agent acts *as if* he has a positive feeling and arrives at an attitude of acceptance of his position of dependence on the trustee for performing X.

2. If additionally the belief sentences of the form bel(A,has_right_to_expect_goodwill(A,B,X)) and bel(*A*,acts_in_part_in_response_to_having_the_right(B,A,X)) were generated a fact of the form normative_trust_performing(A,B,X) will be asserted. For instance if agent 1's belief base contains the following belief sentences pertaining to agent 2

bel(1,intends(2,x)),
bel(1,can_perform(2,x)),
bel(1,goodwill(2,1,x)),
bel(1,has_right_to_expect_goodwill(1,2,x)),
bel(1,acts_in_part_in_response_to_having_the_
     right(2,1,x)),
bel(1,dependent(1,2,x)),
bel(1,bel(2,dependent(1,2,x))),
bel(1,dependence_is_relevant(2,1,x)),

the program rule deduces normative_trust_performing(1,2,x).

3. If only the belief sentences in RSNTR1 to RSNTR5 are present a predicate of the form trust_performing(A,B,X) will be asserted. The generated belief sentences of agent 3

bel(3,intends(2,x)),
bel(3,can_perform(2,x)),
bel(3,goodwill(2,3,x)),
bel(3,dependent(3,2,x)),
bel(3,bel(2,dependent(3,2,x))),
bel(3,dependence_is_relevant(2,3,x)),

match the conditions for trust_performing(3,2,x).

4. If either RSNTR3 and/or a sentence pertaining to trust context Y is lacking, RSNTR4 and RSNTR5

won't be deduced. As argued above the conditions for predictive reliance still hold. For example the following belief base of agent 4 does neither contain a belief sentence representing the goodwill condition nor a belief sentence pertaining to trust context Yc but sentences pertaining to the other conditions exist:

bel(4,intends(2,x)),
bel(4,can_perform(2,x)),
bel(4,dependent(4,2,x)),
bel(4,bel(2,dependent(4,2,x))).

In this case predictive_reliance_performing(4,2,x) will be asserted.

5. No fact will be asserted if RSNTR1 and/or RSNTR2 is lacking.

Every generated trust sentence is logged in the result file. This file can be used as input for a multi-agent system/multi-agent simulation.

Simulation results

During a run the program produces the files beliefs.tru and results.tru. Beliefs.tru contains the agents' belief base, results.tru contains the generated trust sentences for all agents. If an explanation of *ReasonId* was specified, the explanation of the reasons of each agent which lead to his rational social normative trust/ predictive 'trust'/'predictive reliance' are logged in the result file by the predicate trust_reason(*Agent, TrustCondition,ReasonId,Meaning*). The output of the reasons is done in the order of processing. Chained reasons will be printed first. For instance in the following output of a simulation run both agent 5 and agent 8 trust in agent 2, but for different reasons/reason chains.

trust_performing(5,2,x).
trust_performing(8,2,x).
trust_reason(5,rsntr1,r13,b_bel_good_reputation_by_perf_x).
trust_reason(5,rsntr3,r33,b_bel_good_reputation_by_perf_x_
     and_b_lacks_reason_to_act_with_illwill_tow_a).
trust_reason(5,yc,r63,b_free_to_perf_x_and_want_a_matters_
     to_b_due_to_b_want_to_get_good_reputation).
trust_reason(5,rsntr2,r27,b_has_all_the_resources_for_perf_x).
trust_reason(5,ya,r41,a_wants_x_and_b_gets_better_result).
trust_reason(5,yb,r53,b_bel_a_wants_x_and_cannot_perf_x_
     but_b_bel_a_bel_b_can_perf_x).
trust_reason(8,rsntr1,r17,b_is_moral_and_b_bel_he_should_
     perf_x_for_moral_reasons).
trust_reason(8,rsntr3,r37,b_is_moral_and_b_bel_he_should_
     perf_x_for_moral_reasons_and_b_has_no_reason_to_act_
     with_illwill_tow_a).
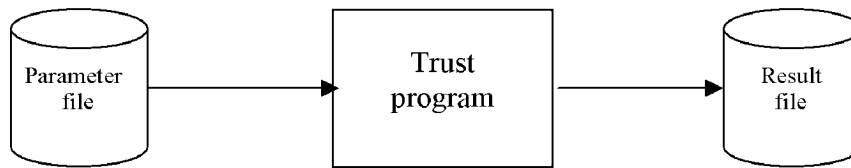
Figure 4. Application as tool for social scientists.

trust_reason(8,rsntr2,r24,b_has_skill_to_perf_x).

trust_reason(8,ya,r43,a_wants_x_and_cannot_perf_x_but_b_
can_perf_x).

trust_reason(8,yb,r53,b_bel_a_wants_x_and_cannot_perf_x_
but_b_bel_a_bel_b_can_perf_x).

trust_reason(8,yc,r61,b_free_to_perf_x_and_want_a_matters_
to_b_due_to_liking).

The results are sensitive to the parameters specified in the parameter file: Giving a high value to *Weight* in the majority of reasons yields a high probability of trust in the artificial society. Vice versa increasing the number of reasons for distrust decreases the probability of trusting agents. The probability of trusting agents is directly proportional to the specified number of agents. If a set of reasons produces a mean of eighty percent of trusting agents in subsequent runs, we get averaged eight trusting agents for a number of ten agents, and a value around eight hundred for a specified number of thousand agents.

The fixed formatted output in results.tru can be used as input for a multi-agent system/multi-agent simulation independently of the programming language of the target system.

Conclusion

Our aim in this paper was to exemplify how criteria for rational trust between agents can be formalized and translated into rules of a programming language. The presented program simulates trust between artificial agents under certain conditions at a single point in time. Although the program rules are very simple the presented simulation program can be used in various ways.

Tool for social scientists

In all cases where trust in societies is an object of study the program can be used as a tool for empirical research. The list of reasons in the parameter file described in section 3 can be customized and adjusted to the object of study. A questionnaire can serve to build the list of reasons for trust and their assigned weights. Subsequent simulation runs will generate different populations and a different number

of trusting/relying agents. By modifying the weights of the reasons probability distributions for trust in the population will be obtained.
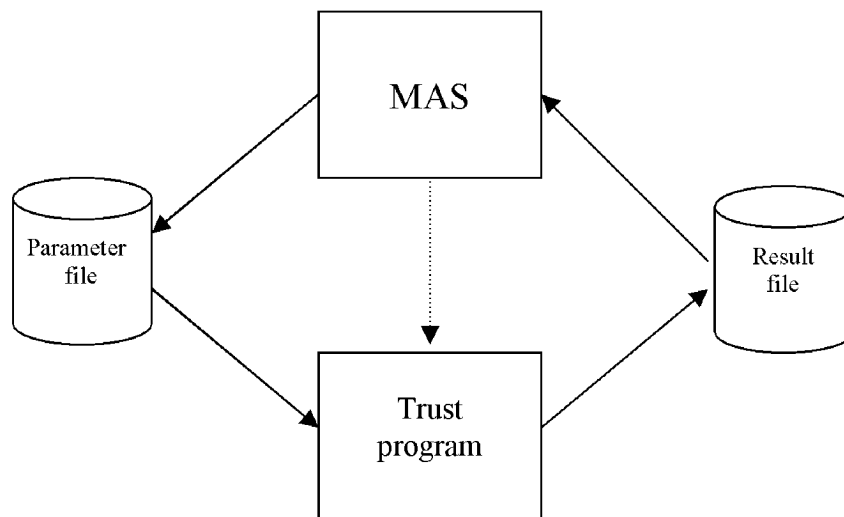
Component of a multi-agent simulation

The program architecture allows a use of the simulation of RSNTR in other multi-agent simulations independently of the programming language. By customizing the parameter file the desired number of agents will be generated. Of course it is reasonable to customize also the reasons given in the parameter file to the object of study. For instance, for a small group simulation which simulates a situation in which the agents know each other, cases which we have labeled friendship, morality, behavior, and group activity can provide reasons to produce the various kinds of trust. In a simulation of negotiations among trading partners who are strangers to each other cases like agreement, reputation or law may be chosen. The main simulation can execute the trust simulation by a system call. After the execution the result file contains sentences expressing which agents have trust in the trustee. These sentences can be used by the decision rules of the main simulation for the courses of action that it suggests.

The target simulation can react to its simulation results by updating the weights of the reasons in the parameter file: If trusting agents were disappointed the weights of the reasons which led to trust should be decreased, if the reasons turned out to be good their assigned weights should be increased.

The trust component could, for instance, be applied to a simulation of some trading partners' behavior in e-commerce in the following way:

- It calculates the probability of the agents' rational social normative trust/predictive 'trust'/ 'predictive reliance' that a transaction will take place under a given set of reasons at a certain point in time.
- By the assumption of 'positive' or 'negative' influences like the presence or absence of fraud and deception a dynamic model can be simulated that shows the change of behavior over time.

Figure 5. Application as component of a multi-agent simulation/multi-agent system.

Component of a multi-agent system

The claims concerning multi-agent simulation hold in principle for other kinds of multi-agent systems as well. For instance, as, contrary to multi-agent simulation, multi-agent systems in electronic commerce often need real-time processing, the program dynamics by file exchange between nodes of the multi-agent system and the example program might be too slow and laborious. In spite of this, we hope to have shown that M. Tuomela's account of trust gives a theoretical basis that can easily be translated into decision rules of a multi-agent system written in any programming language.

Ethical aspects on simulating trust

Simulation as a relatively new method in science cannot be ethical or unethical in itself. In addition to induction and deduction it can be regarded as 'a third way of doing science' (Axelrod 1997). As a 'way of doing thought experiments' it provides an insight into possible conditions/situations. Thus it serves empirical research and theory development as an additional scientific method. Provided that the program is sound and a proper analysis of simulation results is performed, it is an advantage to have the possibility of testing a hypothesis by the aid of additionally generated data. However, if a trust component of the suggested kind would be applied to a 'living' MAS in e-commerce, the issue becomes more difficult. The 'trust' of an artificial agent in another agent can be used as a base for the trusting agent's choice in favor of a certain course of action. Problems can occur if artificial agents make wrong or doubtful decisions. If, for instance, a trading agent starts to order products

from a dubious source, a situation can arise in which it is difficult to determine whether the behavior of the agent is a result of a programming error, of weakly justified probabilistic weights of the reasons supplied in the parameter file, or of a mere subjective mistake. Here 'mistake' refers to the fact that due to the use of random numbers also reasons assigned a low probability can produce the beliefs of trust. Furthermore, even high-probability reasons for forming the trust beliefs may not guarantee truth in all cases. The type of ethical problems referred to above should be dealt with as a part of the ethical problems related to the use of information technology in general and multi-agent systems in particular. It can be argued that artificial agents who have the properties of being interactive, autonomous, and adaptable are moral entities in the sense that they can perform actions to which moral attributes can be meaningfully applied (Floridi and Sanders 2003). However, artificial agents cannot be blamed or punished and they cannot be held morally responsible for their actions. On the other hand it may be difficult to identify the human moral agent who can be assigned responsibility. If we decide to apply systems which are capable of autonomous decision we have to think about the consequences and deal with the ethical questions involved. As all this already holds for systems whose artificial agents may but need not have intentional or mental states, the application of features like a trust component does not add new aspects to the ongoing discussion. The problems discussed by Floridi and Sanders, of course, have to be taken into account for such systems, too. In our case this means that data supplied in the parameter file ought to be designed with special care. Protocols like those given in the result file allow for traceability in cases of error. Still, a great responsibility lies with empirical researchers and with

designers of a multi-agent system. The former have to investigate reasons of human beings that lead to trust, and the latter have to build in and customize the reasons which form the basis of the simulation of trust in an artificial society. When empirical research is used for finding typical reasons for the beliefs of the trust account, the artificial agent can be given an 'attitude' that a typical trusting human agent would have. If the account of trust is a good theory of trust the simulation gives a good base for the agents' choices based on 'trust.'

References

R. Axelrod. Advancing the Art of Simulation in the Social Sciences. In R. Conte, R. Hegselmann and P. Terna, editors, *Simulating Social Phenomena*, pp. 21–40. Springer-Verlag, Berlin Heidelberg New York, 1997.

M. Bacharach and D. Gambetta. Social Trust: A Cognitive Approach. In C. Castelfranchi and Y. Tan, editors, *Trust and Deception in Virtual Societies*, pp. 1–26. Kluwer Academic Publishers, Dordrecht, Boston, and London, 2000.

A. Baier. *Moral Prejudices: Essays on Ethics*. Harvard University Press, MA, 1994.

W. Balzer. SMASS: A Sequential Multi-Agent System for Social Simulation. In R. Suleiman, K.G. Troitzsch and N. Gilbert, editors, *Tools and Techniques for Social Science Simulation*, pp. 65–82. Physica-Verlag, Heidelberg, 1999.

C. Castelfranchi and R. Falcone. Principles of Trust for MAS: Cognitive Anatomy, Social Importance, and Quantification. In *Proceedings of the International Conference on Multi-Agent Systems (ICMAS '98)*. Paris, France, July 1998.

C. Castelfranchi and Y. Tan, editors. *Trust and Deception in Virtual Societies*. Kluwer Academic Publishers, Dordrecht, Boston, and London, 2000.

R. Falcone and C. Castelfranchi. Social Trust: A Cognitive Approach. In C. Castelfranchi and Y. Tan, editors, *Trust and Deception in Virtual Societies*, pp. 55–59. Kluwer Academic Publishers, Dordrecht, Boston, and London, 2000.

L. Floridi and J.W. Sanders. On the Morality of Artificial Agents. Forthcoming in A. Marturano and L. Introna, editors, *Ethics of Virtualities – Essays on the Limits of the Bio-power Technologies*, to be Published for the Series Culture Machine. Athlone Press, London, 2003.

D. Gambetta. Can We Trust Trust. In D. Gambetta, editor, *Trust – Making and Breaking Cooperative Relations*, pp. 213–237. Basil Blackwell, Inc., New York, 1988.

N. Gilbert and K.G. Troitzsch. *Simulation for the Social Scientist*. Open University Press, Buckingham, Philadelphia, 1999.

M. Hollis. *Trust within Reason*. Cambridge University Press, Cambridge, 1988.

S. Hofmann. *Maschinelle Erzeugung und Ableitung von Überzeugungssystemen*. M.A. thesis at the Department of Philosophy, University of Munich, Germany, 1997.

S. Hofmann. The Social Practice of a Women's Group: A First Simulation. In G. Meggle, editor, *Social Facts & Collective Intentionality*, pp. 151–165. Dr. Hänsel-Hohenhausen AG, German Library of Sciences, Frankfurt a. M., 2002.

O. Lagerspetz. *Trust: The Tacit Demand*. Kluwer Academic Publishers, Dordrecht, Boston, and London, 1998.

O. Loukola. Who do you Trust? Combining Morality and Rationality. *Reports from the Department of Philosophy, University of Helsinki*, 4, 1999.

H. Nissenbaum. Securing Trust Online: Wisdom or Oxymoron? *Boston University Law Review*, 81: 101–130, 2001.

P. Pettit. The Cunning of Trust. *Philosophy and Public Affairs*, 24: 202–225, 1995.

T. Scanlon. Promises and Practices. *Philosophy and Public Affairs*, 19: 199–226, 1990.

M. Tuomela. *The Components of Rational Trust and the Power of Social Pressure*. Licentiate thesis at the Department of Social Psychology, University of Helsinki, 2000a.

M. Tuomela. The Mental Components of Rational Trust. In *Proceedings of the Autonomous Agents 2000 Workshop on Fraud, Deception, and Trust, Barcelona, June 2000*, pp. 115–125, 2000b.

M. Tuomela. Rational Trust. In B. Brogaard, editor, *Rationality and Irrationality (Contributions of the Austrian Wittgenstein Society, 8)*, pp. 188–193. Austrian L. Wittgenstein Society, Kirchberg, 2000c.

M. Tuomela. On Rational Trust. In G. Meggle, editor, *Social Facts & Collective Intentionality*, pp. 367–383. Dr. Hänsel-Hohenhausen AG German Library of Sciences, Frankfurt a. M., 2002.

M. Tuomela. The Components of Rational Trust. Forthcoming in M. Sintonen, P. Ylikoski and K. Miller, editors, *Realism in Action: Essays in the Philosophy of the Social Sciences*, pp. 168–180. Kluwer Academic Publishers, Dordrecht, Boston, and London, 2003a.

M. Tuomela. A Collective's Rational Trust in a Collective. *Protosociology, Understanding the Social II – Philosophy of Sociality*, 18–19: 87–126, Vol. 18. www.protosociology.de, 2003b.

R. Tuomela. *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford University Press, Stanford, 1995.

R. Tuomela. *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge University Press, Cambridge, 2002a.

R. Tuomela. Joint Intention and Commitment. In G. Meggle, editor, *Social Facts & Collective Intentionality*, pp. 385–418. Dr. Hänsel-Hohenhausen AG German Library of Sciences, Frankfurt a. M., 2002b.

R. Tuomela. The We-mode and the I-mode. In F. Schmitt, editor, *Socializing Metaphysics: The Nature of Social Reality*, pp. 93–127. Rowman and Littlefield, Lanham, 2002c.

R. Tuomela and M. Bonnevier-Tuomela. Social Norms, Tasks, and Roles. *Reports from the Department of Philosophy, University of Helsinki*, 1, 1992.

R. Tuomela and M. Bonnevier-Tuomela. From Social Imitation to Teamwork. In G. Holmström-Hintikka and R. Tuomela, editors, *Contemporary Action Theory*, Vol. II, Social Action, pp. 1–47. Kluwer Academic Publishers, Dordrecht, Boston, and London, 1997.

R. Tuomela and M. Bonnevier-Tuomela. Norms and Agreement. In E. Attwool and P. Comanducci, editors, *Sources of Law and Legislation*, Vol. III, pp. 87–93. Franz Steiner Verlag, Stuttgart, 1998.

R. Tuomela and M. Tuomela. Acting as a Group Member and Collective Commitment. *Protosociology, Understanding the Social II – Philosophy of Sociality*, 18–19: 7–65, Vol. 18. www.protosociology.de, 2003.