The Knowledge Engineering Review, Vol. 14:1, 1999, 81–89
Printed in the United Kingdom
Copyright © 1999, Cambridge University Press

The challenge of trust, The Autonomous Agents '98 Workshop on Deception, Fraud and Trust in Agent Societies

RINO FALCONE[1] and BABAK SADIGHI FIROZABADI[2]

[1]*Group of "AI, Cognitive Modelling and Interaction", IP-CNR, Rome, Italy. Email: falcone@pscs2.irmkant.rm.cnr.it*
[2]*Department of Computing, Imperial College of Science, Technology and Medicine, University of London, UK. Email: bsf@doc.ic.ac.uk*

1 Introduction

Both the advent of large communication networks and the development of agent models designing more and more sophisticated (in some cases autonomous) behaviours, are producing what can be called an Agent Society. Agent societies are constituted by both artificial and human agents (in perspective indistinguishable from each other) communicating with each other (not necessarily each with all the others), with individual or collective tasks, different resources, different skills, and so on. In other words, these agent societies will become more and more similar to the human ones. As in real societies, in these virtual societies it will be necessary to consider all the problems connected with secure and reliable interaction, with trusting the other agents, with the possibility to be frauded, cheated, deceived in some exchange or relationship.

Humans have developed two main ways to engender trust (Gambetta, 1990). On the one hand, each human being has its own cognitive ability (reasoning, experience, perception, intuition, etc.) to trust the others. On the other hand, at the level of societies many objective techniques (contracts, signatures, long-term personal relationships, etc.) have been developed to ensure trust among agents (and to prevent deception and fraud) with corresponding disciplinary actions to punish the violators.

With the recent research in

- agent theory (in which the notion of agency is studied, and software assistants and mobile agents are intended to act as virtual representatives of (on behalf of) network clients),
- electronic commerce (a new way to establish trading relationships among partners on the Internet),
- human computer interaction (with the relationships between the user and its digital personal assistants), and
- multi-agent systems (for modelling groups, teams, organisations, coordination, negotiation, etc.),

the role of trust is undoubtedly relevant and central for these new agent communities.

In fact, it is necessary to understand what trust is, what its cognitive components are, its social functions. In other words, a principled theory of trust is needed. It is also necessary to identify some automatic techniques and tools to ensure environmental trust (for example, so-called Trusted Third Parties are introduced in an agent society to take care of trust building among the agents in the network). It is important to understand the difference between trusting an agent (as task executor) and trusting the sources (often social) of the beliefs on which the model of the world is based (consider, for example, the case in which there are several information sources, and where it is possible that some of them may not be reliable; making it necessary to characterise the sources that

can be trusted. Finally, it is also important to emphasise the difference between trust and safety: trust building is more than secure communication via electronic networks, as can be obtained with, for example, public key cryptography techniques (the reliability of information about the economical status of a trade partner has very little to do with secure communication with it).

To sum up, a series of issues must be deepened and more clearly analysed: What does it mean to trust an agent? What are the agent's characteristics (cognitive, behavioural, etc.) that permit one to trust that agent? What are the characteristics of a reliable environment (system, society, etc.) independent of the agents? What are the links between the former characteristics and the latter? Are these characteristics general or do they vary according to the task, the domain, the environment? What forms of deception exist? How and why do agents cheat other agents?

Starting from the previous considerations, and from the fact that most agent models neglect all the above-mentioned problems about trust and deception, a workshop on ''Deception, Fraud and Trust in Agent Societies'', organised by Cristiano Castelfranchi, Rino Falcone, Babak Sadighi Firozabadi and Yao-Hua Tan, was held in Minneapolis (on 9 May 1998—in collaboration with the ''International Conference on Autonomous Agents '98'') to collect some significant contributions on this topic, and to better understand the level of maturity of this new area. In the following, we analyse the various papers presented at the workshop. Participants included scientists from computer science, philosophy of right, game theory, cognitive modelling, legal and normative systems, social sciences, multi agent systems, psychology and business systems.

The various contributions—ranging from trust in your counterpart agent not to cheat on you, trust in security of communication infrastructure between agents, to the use of information technology decision support tools to make an expert's opinion more trustworthy—can be described under several headings. According to a content dimension, we have works on deception and works on trust. A second dimension permits us to distinguish between theoretical papers, devoted to considering the foundational aspects of trust and deception, and applicative ones, devoted to identifying the mechanisms for engendering trust in some system. Finally, according to a methodological dimension, we can see several papers in the workshop analysing the notions of deception, fraud and trust using various modal logical operators. Modal logics (Chellas, 1988) are widely applied in theoretical computer science and artificial intelligence. Normally, modal operators are combined to formally represent some notions and also reason about these in certain applications.

2 Trust: Foundational aspects and its formalisation

In his talk ''To trust information sources: a proposal for modal logical framework'', Demolombe discussed trust in the form of an agent's (information source) trustworthiness to deliver correct information. The idea is to give a reasoning mechanism to distinguish between trustworthy and non-trustworthy agents. Demolombe's analysis is based on the following elementary properties: sincerity, credibility, co-operativity and vigilance. The informal definitions for these properties are given as follows:

1 Sincerity: agent A is sincere with respect to agent B when he believes what he says to B.
2 Credibility: agent A is credible when what he believes is true in the world.
3 Co-operativity: agent A is co-operative with respect to agent B if he tells B what he believes.
4 Vigilance: agent A is vigilant with respect to p if, whenever p is true in the world, then A believes p.

The basic modal operators used in this paper are $I_{x,y}p$ reads as ''agent x informs agent y about p'', $B_x p$ reads as ''agent x believes that p'' and $K_x p$ reads as ''agent x strongly believes that p''. The I-operator is, in a sense, a successful operator, since it is assumed that there is no failure in a communication process between two agents. This means that whenever one agent informs another about some fact, then the second agent does not ignore that he is informed about the fact. This is characterised by the following axioms:

$$I_{y,x}p \rightarrow K_x(I_{y,x}p), \text{ and } \neg I_{y,x}p \rightarrow K_x(\neg I_{y,x}p)$$

The I-operator is closed under the inference rule (RE). A normal KD system is defined for the B-operator. The notion of strong belief (the K-operator) is characterised by the property: if agent x strongly believes that p, x strongly believes that, if he strongly believes that p is true then p is true. This is given formally by the following axiom:

$$K_x (K_x p \rightarrow p)$$

Further, the axioms K and D are also adopted for the K-operator. The formal definitions of trust based on the elementary properties mentioned earlier are given as follows:

$$\begin{array}{ll} Tsinc_{x,y}p = K_x (I_{y,x}p \rightarrow B_y p) & \text{(Sincerity trust)} \\ Tcred_{x,y}p = K_x (B_y p \rightarrow p) & \text{(Credibility trust)} \\ Tcoop_{x,y}p = K_x (B_y p \rightarrow I_{y,x}p) & \text{(Co-operativity trust)} \\ Tvigi_{x,y}p = K_x (p \rightarrow B_y p) & \text{(Vigilance trust)} \end{array}$$

The paper contains examples of how one could derive new information from a set of facts and different trust assumptions between the agents.

Some foundational aspects of trust were discussed by Castelfranchi and Falcone in their talk "Social Trust: cognitive anatomy, social importance, quantification and dynamics". The authors tried to individuate the mental ingredients of trust: its specific beliefs and goals, with special attention to evaluations and expectations. They argued the importance of trust in multi-agent systems, showing how cooperation, collaboration, organisations, roles, and so on, are strictly based on some form of delegation of some agent on other agents, and how delegation is strictly based on trust (see Castelfranchi and Falcone (1998)). Delegation is a multi-agent oriented act and relation, which is based on a specific set of beliefs and goals and on a decision: this complex and typical mental state is trust.

Following the authors' thoughts, the basic mental ingredients of trust are:

(i) a goal: the trustor should have a given goal g;
(ii) competence belief: the trustor should believe that the trustee is useful for the goal g, that it can produce the expected result;
(iii) disposition belief: the trustor should believe not only that the trustee is able and can do that action/task, but also that it actually will do that. With cognitive agents this will be a belief relative to their willingness; with objects it will be just a belief about their accessibility and effectiveness.
(iv) dependence belief: the trustor should believe that either it needs a trustee or, at least, that it is better for it to rely on the trustee than not.

When applied to cognitive, intentional agents, the disposition belief must be more articulated:

(v) willingness belief: the trustor believes that the trustee intends to do the action for g;
(vi) persistence belief: the trustor should believe that the trustee is stable enough in its intentions, that has no conflicts with g, or that it is not unpredictable by character, etc.;
(vii) self-confidence belief: the trustor should believe that the trustee knows that it can do some action for g.

For the authors, "to trust" not only means that the trustor has the above beliefs and goals, but also that it decides to delegate the task to the trustee. In other words, the trustor decides not to renounce g, not personally bringing about g, not searching for alternatives to that trustee, and to pursue g through the trustee.

The authors have shown that trust is a bet and implies risk, because it has been derived from its reference to a goal, from the action of delegating, and from the uncertainty of trust-beliefs. Castelfranchi and Falcone showed also how, in their model of trust, the degree of trust is a function of the subjective certainty of the pertinent beliefs: the quantitative dimensions of trust (Deutsch, 1949) are based on the quantitative dimensions of its cognitive constituents. The degree of trust has been used to model the decision on whether or not to delegate.

After this presentation, a wide discussion about the differences between interpersonal and environmental trust was held. The main remark was that people do not trust the agents they delegate trust to; they want to be protected by the contract. The authors' answer was that, in any case, this means dealing with a more complex and specific kind of trust (about contracts, organisations, etc.), but trust remains crucial. This level of trust is a three-party relationship: it is a relation between the trustor, the trustee and an authority. In practice, the trustor trusts the authority and its ability to control, to punish, etc.; the trustor trusts the trustee by believing that it will do what it promised because of its honesty, or because of its respect/fear towards the authority.

In the talk "On the Characterisation of a Trusting Agent—Aspects of a Formal Approach" by Sadighi Firozabadi and Jones, the notion of trust in terms of honesty was discussed and the relation between trust and deception was examined. The authors gave an informal analysis of trust via the notion of deception. The definition of a deceitful action or a lie used in this work is based on that given by Firozabadi, et al. (1998), in which an agent *a* is recognised as a liar if he makes another agent *b* believe something which he (*a*) does not believe himself.

Firozabadi and Jones argue that a successful deception is possible only in the presence of a trust relation between two agents, a liar agent and a victim of the lie. For example, a liar succeeds in his deceptive act only if the victim *b* trusts him with respect to that act. So *a* can attempt to make *b* believe something (i.e. by communicating), but *a*'s success is due to the reasoning mechanism of agent *b*. Here, formal analyses are given to capture such a reasoning mechanism based on trust.

A monadic and dyadic action operator based on Elgesem (1993) are represented; $E_a p$ means that "agent *a* brings it about that p", and $E_a(p,q)$ means that "agent *a* brings it about that p, and this results in that q is true". Both E-operators represent successful actions, since they contain the T-axiom among their set of axioms. The idea of the dyadic action operator is to capture a consequence relation (of some kind) that exists between bringing about that p and bringing about that q.

It is stressed by the authors that an important relation between the two action operators is: $E_a(p,q) \rightarrow E_a q$, which says that "whenever *a* brings it about that p, and this by some consequence relation results in q, then *a* also brings it about that q". Using the dyadic action operator makes it possible to represent a situation in which an agent *a* makes another agent *b* believe that p by, e.g., sending an information denoted by m, as: $E_a(m,B_a p)$. Here, the B operator is a KD45 normal modal operator with the inference rule (RN).

The authors write in their paper that "the essential link between *a*'s bringing about that m, and *b*'s believing p, is provided by the fact that, for *b*, *a*'s bringing about that m counts as sufficient evidence for p". This fact is represented using the relativised conditional connective (in the style of Jones and Sergot (1996)): $(E_a m \Rightarrow_b p)$, which represents that *b* trusts that *a*'s bringing about m indicates that p holds. This means that the same message sent by agent *a* to another agent *c* may not result in that *c* believes p holds; or that the same message sent by agent *c* to agent *b* may not result in that agent *b* believes p holds. Given these operators, the authors give a new representation of a deceitful action as follows:

$$\neg B_a p \text{ AND } E_a B_b E_a m \text{ AND } B_a(E_a m \Rightarrow_b p)$$

which says that *a* does not believe that p, and that *a* brings it about that *b* believes that *a* brings it about that m, and that *a* believes that his bringing about that m counts for *b* as sufficient evidence for the truth of *p*. By this definition, the authors capture the idea that *a*'s success in his deceitful action is due to the fact that *b* trusts him, and takes his sending the message m as a sufficient evidence for the fact that p holds. So, in case agent *a* is mistaken in his belief that *b* trusts him, then *a* will fail in his attempt to deceive agent *b*.

3 How to realise a trusted environment: towards electronic commerce

Another talk based on formal analysis using modal operators was "A formal specification of automated auditing of trustworthy trade procedures for open electronic commerce" by Bons, Dignum, Lee and Tan. The authors argued the importance of such procedures in open electronic

commerce, where there is a lack of trust between trading parties. Examples of such trade procedures are documentary credit procedures, in which a set of documents is exchanged between trading parties to control each others actions.

This work concentrates on modelling the design principles given by Bons (1997) for designing trustworthy trade procedures. These principles are rule bases using the audit daemons formalism developed by Bons (1997). There are three steps involved here: (1) it is determined whether the right documents are exchanged among the right parties; (2) the technical documents that should be satisfied by the underlying communication system used to exchange the documents are identified; and (3) whether the actual designed systems do satisfy these requirements is analysed.

The components of the trade procedures that have to be modelled are identified as (1) the trade procedure itself, (2) the underlying contracts, and (3) the audit rules that check the trustworthiness of the procedure, given the underlying contracts. The language for modelling the above components is suggested to be based on various modal operators that can capture the following notions: obligation, action, temporal relation, knowledge or belief, delegation and trust. It is acknowledged that a Prolog-based implementation of the audit daemons, called INTERPROCS (Lee, 1988), does already exist. However, the authors argue that using modal logic would give a richer formal specification language. A multi-modal logical language is presented which contains operators to express the above-mentioned notions. Some examples of how one could model and also check a trade procedure according to the specified design principles using the suggested specification language are also given. In this language, each action in a trade procedure is specified as a predicate. A trust relation between two agents is also given in terms of a predicate: Trust (a,b,p), which says that *a* trusts *b* with respect to the performance of p, where *a* and *b* represent two trading parties and p represents a certain activity.

He and Sycara, in their talk ''Towards a Secure Agent Society'', focused their attention on what a ''secure agent society'' should be and how to develop it. In their view, this security consists of three aspects:

(i) agent authentication mechanisms that form the secure society's foundation;
(ii) a security architecture design within an agent that enables security policy making, security protocol generation and security operation execution;
(iii) the extension of agent communication languages for agent secure communication and trust management.

The authors think that agents, as primarily human-delegated software, could become the primary area in which to practice modern cryptographic technology, and this for at least two reasons: first, the software agents are playing an increasingly active and bigger role in electronic commerce; secondly, the execution of most security protocols would be a big burden for the end-user, but it could be done fairly easily by autonomous software agents.

Referring to RETSINA (Sycara et al., 1996), an intelligent multi-agent project at Carnegie Mellon University, the authors explain how they are implementing an agent-based Public Key Infrastructure (PKI) (Polt and Dodson, 1998). Unlike traditional PKI implementation, their PKI implements the authorities of authentication verification service systems as autonomous software agents, called security agents, instead of building a static monolithic hierarchy. From the user viewpoint, the security agent can be thought of as a kind of configurable facilitator that can be employed by any group of agents, or the owner of the agents, to construct their own authentication verification service system. ''Configurable facilitators'' do not pre-specify any particular certification format or hierarchical relationship in the software (as in other traditional PKI projects), but allow users to define the format(s) of the certification(s) and the name space(s) as they need (customising).

In RETSINA there are three modules involved in agent security, AgentEditor, Planner and Security Module, which correspond to policy specification, protocol generation and operation execution. Defining a set of security policies for a given task is the first level job for agent security, and it would be done by the owner of the agent. According to the security policy, a security protocol

must be generated by a functional module (planner) as part of a procedure for the agent to complete the task. To execute the security protocol, some basic security functions, such as encryption, decryption, signing, verification, etc., would be called during the execution of task. Finally, the authors introduced an extension of KQML (Knowledge Query and Manipulation Language—a communication language for agents) for agent security, including a set of new parameters and a set of new performatives.

In his talk "Engendering Trust in Electronic Environments—Roles for a Trusted Third Party", Rea showed the initial results of a project investigating the nature and role of trust in the online world. The author considered how, within the electronic domain, the concept of a Trusted Third Party (TTP) has evolved out of the development of public key infrastructures. Until now, the term "trust services" has often been applied to a set of functions based upon the use of public key cryptography to ensure confidentiality, and to create "digital signatures" for identifying verification and to ensure message integrity. These cryptographic techniques rely upon the presence of a TTP, which enables users to verify the validity of the keys they are using to encrypt and decrypt information for and from other parties.

The author declares that the technology of trust services will be an important enabler for electronic commerce, but to begin to realise the full potential of electronic commerce it will be necessary to understand how trust services can be used to support trust, and the TTP role should be extended.

Rea individuates two main roles for the TTP: the Environmental Role (ER) and the Transactional Role (TR). In the ER, a carefully defined and controlled environment with a controlled user community will have a greater sense of shared values, and hence a higher level of environmental trust, than an open community with a more flexible membership. The TTP can vet users prior to membership to ensure that they meet certain minimum criteria. This will allow users of the system to make basic valid assumptions about the trustworthiness of all other users. The TTPs role in engendering transactional trust will—following the author's thought—be based upon providing functionality for allowing potential trustees (the term *trustor* is used by Rea to refer to the actors who are evaluating trustworthiness, and *trustee* refers to the actors who are being evaluated) to communicate their trustworthiness and functionality for allowing potential trustors to evaluate trustworthiness.

In terms of risk control, the TTP can act to structure transactions such that only a minimal level of trust is required to enter into the transaction. Finally, the author says that the TTP can implement its trust-engendering and risk-controlling functions through the careful design of the agents upon which the trading environment is based.

In the first part of the paper "The Role of Trust in Designing Trade procedures", Ganzaroli, Lee and Firozabadi discuss the relation between trust and control in a trade procedure. Here, the authors argue that the level of control applied by trading parties in a trade procedure corresponds to the level of trust between the parties. In other words, if the trading parties trust each other completely, then there is no need for control routines, but they are needed when there is a lack of complete trust between the agents. The authors define the level of controls in a hierarchical structure, such that for each control routine there is a need for higher level control that checks the validity of the lower level activity, but only if the parties do not trust each other completely or if the risk involved is negligible. The argument is that trust is needed to stop a regression of control levels.

In the second part of the paper, the authors give an overview of research in progress. The core of this research is to develop a computational formalism called Procedure Constraint Grammars, which can be used to combine different trade procedures in order to design a new procedure. The system should give different alternatives of trade procedures based on the old procedures and the new trust relations between the parties, so that the parties will be able to determine and choose the most trustworthy procedure from among the alternatives.

In their talk "Trust in Electronic Commerce—a Language/Action Perspective", Weigand and van den Heuvel discussed the notion of trust in a communication, in the tradition of Winograd and

Flores (1986) and Habermas (1984). It is argued that the core of electronic commerce transactions is the communicative acts performed by the parties in the form of messages and documents. A taxonomy of social actions is given based on Winograd and Flores (1986). The social actions are classified into two main classes: the strategic actions and the communicative actions. The strategic action of a party is based on his opportunistic behaviour in a transaction to achieve individual goals. The communicative action is oriented towards mutual understanding as a way of co-ordinating the actions of the parties. Trust is then defined in terms of these two types of actions as one party's assumption that the other party will not act strategically, but is acting communicatively. The issue of security is discussed in the paper, and the authors argue that security measures are needed both for the means and also at the ends of communication. The authors argue that in their framework, ''electronic commerce'' is doing business by means of speech acts (Austin, 1962; Searle, 1969), hence security should be applied both at the utterance acts (the physical level) and also the illocutionary acts (the essential level). In the same way as in Bons (1997), Weigand and van den Heuvel distinguish between the communication level and the business level. The issue of security is discussed in these two levels in terms of speech acts.

The second part of this work is concerned with the role of a broker in facilitating communication between parties that want to engage in a business transaction. Their argument is that such a broker has to be trusted by the parties in order to reduce the transaction costs. The authors argue that the trusted broker should provide security both at the communication and also at the business level.

4 Exploring deception

To investigate the role that deception can play in multiagent worlds, and the forms of deception that agents can apply in the different circumstances, Castelfranchi, Falcone and de Rosis discussed their work, ''Deceiving in GOLEM: how to strategically pilfer help''. GOLEM (Castelfranchi et al., 1997a) is a system aimed at formalising, implementing and experimenting on different kinds of levels of social cooperation—they are studying both different social attitudes and personalities in interaction (in particular, attitudes in delegating tasks to other agents and in adopting the goals of others), and the reasoning about delegating and adopting tasks, and about the other's personality (Castelfranchi et al., 1997b).

The authors classified several forms of deception (all allowed in GOLEM), for example:

- *About the agents and about the ''physical'' world*: because agents have different views of the world-domain (there are parts of the world that they cannot see, while other agents can), any deception about the world is possible.
- *Intentional, and accidental deception*: GOLEM agents are intentional and planning agents, thus, that other agents believe or not believe something may be among their intentions; GOLEM agents, however, do not know (and then intend) all the effects of their actions, so they might accidentally deceive others.
- *Passive and active deception*: it is a passive deception when the deceiving agent simply lets the other agent ignore something crucial for him, or believe something to be wrong.
- *Deception with communication and without any communication*.
- *Deception by implicit, behavioural communication (bluff) and by explicit and specialised communication acts (lie)*.

But why to deceive others? In GOLEM the deception is only instrumental to other goals: it is just due to conflicts: conflicts between the domain plan/goal of the agents (if an agent achieves its goal, the other agent cannot achieve its own goal); conflicts between the social goals. Finally, in their analysis the authors showed how to deceive; the agents must have an image of the other agent's mind and be able to reason on it, in several ways (considering capabilities, personalities, goals, plans, etc.).

5 An exploratory study of trust

In his talk ''Developing Trust with Intelligent Agents: an Exploratory Study'', Elofson tried to cover several areas: definition of trust, components of it, the role of information technology in trust, empirical evaluation of trust. Several perspectives on the meaning of trust were examined and then synthesised to form a new (composite) definition of trust: *trust is the outcome of observations leading to the belief that the actions of another may be relied upon, without explicit guarantee, to achieve a goal in a risky situation*. Whether or not the exact nature of those actions can be enumerated is unspecified in this definition of trust. The expected actions may be known or unknown.

Referring to McCroskey (1966), Elofson thinks that trust is made up of both a character-based component and an authority-based component. The focus of his presentation was on this second component, with its strong implications for competence, credibility, expertise, experience and professionalism. In particular, *professional trust* must be created and maintained. But how? How does one engender the sense that another individual, because of their expertise, their competence, their experience and their credibility, may be relied upon to achieve a goal in a risky situation? Starting from the ideas of Butler (1983), Quinones (1995), Rosman et al. (1994) and Jennings (1967), Elofson concluded that predictability, neutrality, the use of procedural knowledge and openess prefigure the creation of professional trust by fostering a belief in an individual's expertise, experience, competence and credibility. In particular, the author thinks that intelligent agents—acquiring the heuristics that decision makers use over time and assuming responsibility for a growing number of decision-making activities proportional to the tasks given to them—may play a vital role in the promotion of trust in organisations. Intelligent software agents may generate a series of expert-system like if/then rules. The availability of the procedural knowledge encoded in the rules, together with their logical nature and the explicit way in which they detail an individual's decision process, should significantly increase the expression of credibility, experience, confidence and predictability that the heuristics convey.

Finally, Elofson described his empirical studies in which, using an existing intelligent agent system (Knowledge Cache—a system for supporting the applied task of environmental scanning), he proves the positive effects of an intelligent agent-generated set of decision heuristics on trust. In these experiments it is the information pertaining to the problem-solver, rather than the problem, that affects trust: intelligent agent-generated decision heuristics can aid in the development of professional trust.

6 Conclusions

Considering the few papers presented on deception and the discussions on this theme in the workshop, it seems clear that deception is mostly considered as a sort of game topic. The relevance and role of this issue has not yet been pointed out in agent communities.

On the contrary, the wide, heterogeneous and crowded participation to the workshop on the theme of trust testifies to the increasing importance of this topic. It will soon become an important theme of research for several areas. All this interest is strictly linked with a real emerging problem: is it possible that an advanced technology (such as the Internet and the automatic processes on it) may be stopped by a psychological issue like the need for trust (for example, electronic commerce cannot work without trust)? In other words and more generally, is it always necessary for technology to consider the human factors? Are they so crucial for a suited and useful development? The answers coming from this workshop are all pointing to the need for this strong relation.

Acknowledgements

We would like to thank Cristiano Castelfranchi and Yao-Hua Tan (chairs of the workshop) for their useful comments to this paper.

References

Austin, J, 1962. *How to do Things with Words* Clarendon Press.

Bons, RWH, 1997. "Designing trustworthy trade procedures for open electronic commerce" PhD thesis, Erasmus University, Rotterdam.

Butler, JK, Jr, 1983. "Reciprocity of trust between professionals and their secretaries" *Psychological Reports* **53** 411–416.

Castelfranchi, C, de Rosis, F, Falcone, R and Pizzutilo, S, 1997a. "A testbed for investigating personality-based multiagent cooperation" *Proc. Symposium on Logical Approaches to Agent Modeling and Design* Aix-en-Provence, August, pp. 18–22.

Castelfranchi, C, de Rosis F and Falcone, R, 1997b. "Social attitudes and personalities in agents, socially intelligent agents" *AAAI Fall Symposium Series 1997* MIT, pp. 16–21.

Castelfranchi, C and Falcone, R, 1998. "Principles of trust for MAS: cognitive anatomy, social importance, and quantification" *Proc. International Conference of Multi-Agent Systems (ICMAS'98)* pp. 72–79.

Chellas, BF, 1988. *Modal logic: An Introduction* Cambridge University Press.

Deutsch, M, 1949. "Theory of cooperation and competition" *Human Relations* **2**(2) 129–152.

Elgesem, D, 1993. "Action theory and modal logic" PhD Thesis, Institutt for Filosofi, Universitetet i Oslo.

Gambetta, D (ed), 1990. *Trust* Basil Blackwell.

Habermas, J, 1984. *The Theory of Communicative Action: Reason and Rationalization of Society, vol. 1* Beacon Press.

Jennings, EE, 1967. *The Mobile Manager* University of Michigan.

Jones, AJ and Sergot, MJ, 1996. "A formal characterisation of institutionalised power" *J. IGPL* **4**(3) 427–443.

Lee, R, 1988. "INTERPROCS: A Java-based prototyping environment for distributed electronic trade procedures" *Proc. 31st Hawaii International Conference on System Sciences (HICSS'98)* IEEE.

McCroskey, JC, 1966. "Scales for the measurement of ethos" *Speech Monography* **33** 65–72.

Polk, WT and Dodson, DF, 1998. "Public Key Infrastructure: From Theory to Implementation" `http://csrc.ncsl.nist.gov/pki/panel/overview.html NIST`.

Quinones, MA, Ford, JK and Teachout, MS, 1995. "The relationship between work experience and job performance: A conceptual and meta-analytic review" *Personnel Psychology* **48**(4).

Rosman, A, Lubatkin, M and O'Neill, H, 1994. "Rigidity in decision behaviors: a within-subject test of information acquisition using strategic and financial informational cues" *Academy of Management J.* **37**(4).

Sadighi Firozabadi, B, Lee, RM and Tan, Y-H, 1998. "Formal definitions of fraud" in McNamara, M and Prakken, H (eds), *Proc. Fourth International Workshop on Deontic Logic in Computer Science* Bologna.

Searle, J, 1969. *An essay in the philosophy of language* Cambridge University Press.

Sycara, K, Decker, K, Pannu, A, Williamson, M and Zeng, D, 1996. "Distributed intelligent agents" *IEEE Expert* 36–45.

Winograd, T and Flores, F, 1986. *Understanding Computers and Cognition: A New Foundation for Design* Addison-Wesley.