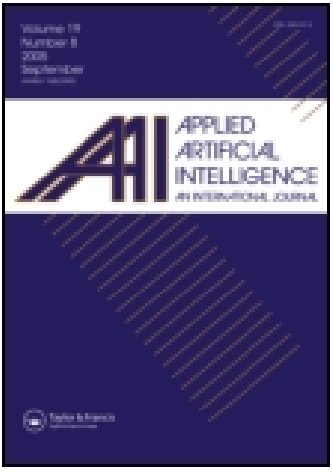

Applied Artificial Intelligence: An International Journal

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uaai20

Guest editorial

Cristiano Castelfranchi [a] , Rino Falcone [a] , Babak Sadighi Firozabadi [b] & Yao-Hua Tan [c]

[a] Division of "Artificial Intelligence, Cognitive Modeling and Interaction" , National Research Council, Institute of Psychology , Rome, Italy

[b] Swedish Institute of Computer Science, Kista, Sweden, and Department of Computing, Imperial College , University of London , United Kingdom

[c] Erasmus Center for Electronic Commerce (ECEC) , Erasmus University Rotterdam , Rotterdam, The Netherlands

Published online: 26 Nov 2010.

To cite this article: Cristiano Castelfranchi , Rino Falcone , Babak Sadighi Firozabadi & Yao-Hua Tan (2000) Guest editorial, Applied Artificial Intelligence: An International Journal, 14:8, 763-768, DOI: 10.1080/08839510050127533

To link to this article: http://dx.doi.org/10.1080/08839510050127533

Applied Artificial Intelligence, 14:763–768, 2000

0883-9514/00 $12.00

GUEST EDITORIAL

CRISTIANO CASTELFRANCHI and
RINO FALCONE[1]
National Research Council, Institute of Psychology, Division of "Artificial Intelligence, Cognitive Modeling and Interaction", Rome, Italy

BABAK SADIGHI FIROZABADI
Swedish Institute of Computer Science, Kista, Sweden, and Department of Computing, Imperial College, University of London, United Kingdom

YAO-HUA TAN
Erasmus Center for Electronic Commerce (ECEC), Erasmus University Rotterdam, Rotterdam, The Netherlands

WHY TRUST?

Trust and related issues such as reputation, source reliability, and deception are becoming really hot topics in information technology (IT), and in particular in artificial intelligence (AI) domains like Agents and Multi Agent Systems (MAS).

Why is trust so important? These are some generic reasons:

- the increasing relevance of security, safety and privacy issues – and of their perceived or subjective counterpart—in IT and in particular on the web;
- the fact that trust is essential for any organization, or any human social relation, and thus should be maintained and created in new forms of computer supported collaboration and computer mediated communities;
- the fact that what is growing on the web is a new market, Electronic Commerce (EC), and since a market is mainly based on selfish interests, it is open to cheating and to competition between attractive alternatives. Hence, this new virtual market strongly requires trust in the chosen partner, trust in rules and procedures, trust in guaranties and enforcing authorities.

Trust is also important for some specific features of new environments and technologies, such as relations with unknown persons, across different countries, cultures and laws; the possibility of anonymity and of changing

identity; the non face-to-face cooperation (where defection seems more tempting); the non-filtered and guaranteed information one can find on the Web (rather different from printed information where there are trust-creating institutions such as publishers, editors, journals and newspapers, and more clear norms, conventions and responsibilities).

Trust is also important for something intrinsic to the new computational paradigm which is open, distributed, dynamic and based on some *autonomy*. The true basis of trust—as Luhman explained—is uncertainty: in order to deal with an uncertain world and to decide to act and pursue our goal without perfect knowledge and stable environment, we have to take risks, we have to trust enough our information and beliefs, our action, our supports, and other agents we are relying upon to fulfil our needs.

WHY IN PARTICULAR 'TRUST IN AGENTS'?

There is *a peculiar relation between agents and trust.*

On the one hand, Agents are software entities (or robots) acting on the behalf of" a user or designer, i.e. to satisfy the user's requests or needs (this is the agent's task or mission). Notice that this entails that someone is relying upon the action of the agent to satisfy its own goal; it is delegating this to the agent. This cannot be done without some trust. Moreover, since agents can allocate tasks or sub-tasks to other agents, exchange and cooperate, the same trust relations are needed among them. On the other hand, Agents are peculiar software entities that act in an *autonomous* way, i.e. they act "without the direct and complete intervention or control" of the user. More precisely an agent can independently access information, react to local contexts and events, have an evolving internal state that changes its processing, learn or evolve, and can "decide" *how* to achieve a non-completely specified task and *with whom.*

This is true not only for agents on the web or in hybrid and open Multi-Agent Systems MAS, but also in Human Computer Interaction (HCI) where the interactive approaches can entail more initiatives also on the side of the machine. In sum, by 'Trust in Agents' we mean both trust placed in an agent and in agent-based systems by some user, and trust among agents in MAS or in agent to agent interactions (and more broadly and generally the problem of trust in distributed computing and infosociety).

How can we represent this trust, build and maintain it? How to use trust, reputation, and so on, in electronic commerce, in information seeking, in virtual organizations, in electronic communities, etc. How to deal with deception, fraud, malicious intentions, incompetent partners or information sources? Which is the relationship between trust and security, trust and dependability, trust and efficiency? This special issue is dedicated to some

theoretical and practical attempts to answer some of these questions. It is the result of an international workshop on these topics.[2]

DIFFERENT TYPES AND APPROACHES

As the reader will see there are various kinds or facets of the phenomenon of Trust that can be modeled also in a quite independent way. In particular, let us stress the most relevant distinctions, for example between

- *trust in information sources* and information credibility, and
- *trust in a 'partner'* which is expected to perform some desirable action.

Another very important distinction is between

- trust in the *internal* characteristics of an agent that make it trustworthy (and in the signs of this)
- trust in the *external* conditions that can make the action successful or reduce the probability or the damage of its failure. For example, trust in technology, in organization, in rules, protocols and procedure, mediators, assurances, norms, authorities.

Several other distinctions are of course possible and useful, and some of them are in fact introduced in the papers.

Not only are there different types or facets of trust, but there are also different theoretical and practical approaches for dealing with trust.

The main distinction is between a more cognitive or psychological approach aimed at characterizing trust basically as expectations, beliefs, desires, attitudes, feelings or whatever, thus modeling mind or personality (this is typically a qualitative approach, which sometimes also uses some quantitative measures), and other approaches that are basically aimed at operationalizing this notion in economics or in applications. These approaches are mostly quantitative. Within these approaches particularly important is the game-theoretic one, that basically defines trust as the subjective probability of a desirable action. Other approaches reducing trust to some measurable index are used for modeling learning trustworthiness of others through repeated interactions or the dynamics of trust on the basis of experience and sources, and so on. Yet another approach is aimed at formally modeling trust based on some modal logic that can represent basic notions such as commitments, norms, reliability and contracts.

All those approaches are relevant and useful for progress. They compete with each other but they are in fact also cooperating in a long run enterprise about a very rich and complex phenomenon that must be modeled from different perspectives.

IN THIS ISSUE

In his article "Boosting Cooperation By Evolving Trust," Andreas Birk models trust as an emergent property in a complex dynamic system. He distinguishes between trust and trustworthiness (which is defined as an intrinsic property of an individual i_A in respect to another individual i_B). Birk assumes that the trustworthiness is an objective criterion in the sense that it gives i_B a measure allowing a rational choice of whether to interact with i_A or not. The problem is that the trustworthiness is not perceivable in the general case (neither by i_B nor by i_A itself).

Given the opacity of the trustworthiness of an agent, to built trust means to approximate the agent's trustworthiness through some process including an interaction procedure with it. In Birk's article, trust is established through the preferences of agents to be grouped together with other agents carrying a certain marker. Groups play a game based on an extended version of Prisoner's Dilemma. Strategies of the agents in the iterated game establish their trustworthiness. Agents update their trust on the basis of limited interactions with other agents. In the experiments reported in this article, there is neither a correlation between labels and strategies, nor between preferences and strategies in the beginning of each experiment. Nevertheless, stable relations of trust emerge. In addition, it seems that a higher cooperative level in the population is reached faster with the trust building than without it.

In their article "Limiting deception in groups of social agents," Anish Biswas, Sandip Sen, and Sandip Debnath investigate two mechanisms to limit the exploitation of the reciprocative strategy by deceptive agents: 1) a penalty factor for declining requests for help, and 2) a cut-off limit on outstanding balance of help. They evaluate the relative effectiveness of these mechanisms for augmenting robustness of agent behaviors.

The authors find interesting results showing that the first mechanism adds a penalty factor to the mutual balance between agents to decrease the probability of helping an agent who has declined a request. They also show how with a reasonable choice of parameters, both these mechanisms significantly improve resistance to exploitation without noticeably decreasing cooperation potential between similar agents.

In their paper, "Trust and Control: A dialectic link" Cristiano Castelfranchi and Rino Falcone analyze the complex relationship between trust and control. They show how, on the one hand, it is true that where and when there is trust there is no control, and vice versa. But also that this is a restricted notion of trust: it is "trust *in* the trustee", which is just a part, a component of the complete trust needed for relying on the action of another agent. Thus they claim that control is antagonistic to this strict form of trust; but also that it completes and complements it for obtaining a *global*

trust. In other words, establishing control and guarantees is trust-building; it produces a sufficient trust, when trust in trustee's autonomous willingness and competence would not be enough.

They also argue that control requires new forms of trust: trust in the control itself or in the controller, trust in y as for being monitored and controlled; trust in possible authorities; etc.

Finally, they show that paradoxically control could not be antagonistic of strict trust *in* the trustee, but it could even create, increase the trust in it, making the trustee more willing or more effective.

In their article "Using Trust for detecting deceitful agents in artificial societies," Michael Schillo, Petra Funk and Michael Rovatsos propose a *model of trust* that is established by an observation and communication process.

Agents start out with no knowledge about the behavior of other members in the society and then modify their model on trustworthiness of others according to observations and testimonies from agents that witnessed interaction behavior. While interacting and observing, the model about other agents is refined and used to judge their reliability to commitments about future actions.

They model egoistic and altruistic personality profiles, but they provide them with a fuzzy factor: each agent plays according to its social attitude with a given probability. An extension of the Prisoner's dilemma game enhanced with a partner selection phase (contract-net like protocol) is applied.

The authors found that after a number of rounds deceiving agents are excluded from playing because they are no longer trusted by the others.

In their article "A generic model of trust for electronic commerce" Yao-Hua Tan and Walter Thoen present a generic model of trust for electronic commerce. The basic idea of the model is that an individual will only engage in a transaction if his level of trust exceeds his personal threshold, which depends on the type of transaction and other parties involved in the transaction. They argue that the two basic components of the level of transaction trust are the trust in the other party and the trust in the control mechanisms, and they explain that both kinds of trust have objective and subjective aspects. They argue that the generic trust model can be used for the design of trust related value-added services in electronic commerce: they discuss two activities in electronic commerce that require trust, namely electronic payment and cross-border electronic trade. They show with their model how these two activities actually require two different types of trust: trust in international business to business electronic trade is primarily created by an information service, whereas trust in electronic payment systems is created by massive adoption of these systems by trusted companies.

NOTES

1. The initiative of this Special Issue and our contribution to it has been developed within the European Project ALFEBIITE (A Logical Framework for Ethical Behaviour between Infohabitants in the Information Trading Economy of the Universal Information Ecosystem): IST-1999-10298.
2. Autonomous Agents 1999 Workshop in "Fraud, Trust and Deception in Agent Societies", Seattle, May 1.