The Knowledge Engineering Review, Vol. 14:2, 1999, 109–118
Printed in the United Kingdom

Artificial *social* intelligence: a necessity for agent systems' developments

ROSARIA CONTE

Division of AI, Cognitive and Interaction Modelling – PSS (Project on Social Simulation), IP/CNR, V.LE Marx 15-00137 Roma, Italy. Email: rosaria@pscs2.irmkant.rm.cnr.it

1 Introduction and summary

Often, the problems encountered by applications and developments of software agents require further investigation and improvements of agents' social capacities, that is, of the properties enabling them to act adaptively in multiagent domains. This is so for at least three reasons. First, quite often software agents are designed for reasoning upon and interacting with other agents (both human and artificial), as is the case in client-server and personal assistants. Secondly, to accomplish their primary task, these systems ought to reason upon, predict and eventually coordinate with, or at least prevent obstacles from, other systems accomplishing the same or different tasks in a *concurrent* way (think of electronic commerce, meeting agents, etc.). Thirdly, many applications are actually distributed over a set of agents with their specific roles and tasks, which are expected to participate in a *joint* activity, aimed at achieving a given global outcome.

Several kinds of well-known concerns/expectations run throughout the field of software agents relative to their actual applicability: effective performances and reduced errors, greater reliability and robustness combined with higher flexibility and adaptability, context-sensitivity and situatedness, etc. Besides, software agents are situated in increasingly *complex* social environments and settings. This growing complexity is generally environmental, and specifically social: agent systems facilitate the management of dislocated or cross-organisations, as well as the functioning of enlarging networks, virtual worlds and settings, etc.

The recent scientific events in the area of software agents show a general tendency towards applying software agents endowed with social intelligence, i.e., with the capacity not only to *act in* the presence of and *with* others, but also to reason and *act upon what* is known of others, and more precisely, what is known of their mental states. Two specific predictions seem justified:

- Developments in the area of client-server interaction imply an increasing capacity of agent systems to provide some form and degree of *overhelp*, in order to assist their users in an intelligent and effective way.
- Whether agent systems are designed for negotiation or joint activity (i.e., whether they operate in competitive or cooperative contexts), in dyadic or collective interaction they will need to master not only *inter*-agent, but also *multi*-agent, action and coordination. More specifically, they will need both micro- and macro-social abilities, attitudes and knowledge.

This will give further impulse to multi-agent systems research, and will encourage a decisive investment of scientific and interdisciplinary resources in the field of Social Artificial Intelligence (see also Malsh & Schulz-Schaeffer, 1996).

The remainder of the paper is organised as follows. The next section presents current tendencies in the field of agents as emerging from outstanding and recent events (e.g., the Autonomous Agents Conference, the Workshop on Agent Theory, Language and Architecture, etc.). In section 3, some fundamental problems posed by these emerging trends will be identified. It will be argued that to solve them requires further theoretical investigation and consequent implementation in the area of

agent (micro- and macro-)social capacities. In section 4, possible misunderstandings about Artificial Social Intelligence will be discussed, and some conceptual claims will be made. In the concluding remarks, further requirements for development in this field will be suggested.

2 Reports from the agent systems world

Recent scientific events in the area of software agents show the following tendencies:

1. From reaction to preposition (cf. Goldman & Baral, 1998): systems not only need to react to events, but also to plan (or pre-plan) in order to make reaction feasible. This leads, on the one hand, to enabling agents to simulate events and actions in a represented world, and on the other manipulate their own internal states. These developments are facilitated by the introduction of softbots (Etzioni, 1993), and especially of what are called *immobots* (Williams & Nayak, 1996), i.e., autonomous agents dealing more with internal states than with physical action.
2. From support to substitution: expert systems assisting humans in decision-making increasingly act on behalf of humans, and play social (mediator and management) roles. Current generation assistants "guide" their human counterparts and are "guided" by them (Rao, 1998, p. 176).
3. From assistance to overhelp: software agents used as expert assistants (in entertainment, education, tutoring, etc.) tend to be endowed with a personalised knowledge of their users' goals, interests, styles/strategies, preferences, commitments, etc. As is well-known, this raises the issue of trust, since agent systems are allowed to access private if not intimate information about users. Besides, on the grounds of this pre-existing or current knowledge, personal assistants are expected not only to (a) provide their users with a personalised assistance (for example, to filter information), but also to (b) act on their behalf (for example, find information and negotiate for goods and services for them), and even to (c) protect (*to look over their shoulders*; cf. Crabtree, 1998) and influence their users (for example, monitoring how they performed their tasks, drawing their attention to important events and news, etc.).
4. From support to sharing: when used as assistants, believable agents are designed to express emotions associated with given perceptions and beliefs (rather than to simply convey information).
5. From negotiation to partnership: a mediator or representative agent is not only held to look after its user's interests, but also to be accepted by the community into which it will interact. Indeed, Bargainfinder "was one early agent that managed to get banned from a number of CD stores because its aims were possibly not beneficial to any particular store" (cf. Crabtree, 1998, p. 135). More recently, agent systems for negotiation have been developed to deal with this problem (Gutmann et al., 1998).

These shifts justify Rao's prediction (1998, p. 177), that the development of software agents will stimulate work in the areas of believable agents and autonomous robots, and in the area of knowledge-based assistants. Here, it will be argued that developments in both areas are also expected to stimulate work on other aspects of agent systems, namely on their micro- and macro-social capacities. Micro-capacities include overhelp (to know and reason about the users' mental states, to act in their interests and on their minds). Macro-capacities allow agents to act as partners in interaction and members in coalitions.

3 Agent systems: hits and misses

Let us examine two main directions of development: expert assistants and applications to multiagent domains (whether competitive or cooperative).

3.1 Expert assistants

As seen above, expert assistants are increasingly used to play two important social roles:

1. To *represent* their users, i.e., to act on their behalf.
2. To *monitor*/*control* their users.

3.1.1 Expert representatives

The development of personal representatives has several consequences, and raises some important scientific problems.

- *Trust and overhelp*. Often "users trust their agent to work on their behalf" (Crabtree, 1998, p. 135). To work on the users' behalf, agents must provide some sort of overhelp: they must go beyond the users' explicit requests and take decisions which concern other needs, goals and interests of the users. More specifically, these include (a) the current goals of the users that are not explicitly related to the delegated task but with which this task might interfere; (b) potential goals of the users that may reasonably be derived from the dynamics of personal traits (for example, one may not want to enjoy a good reputation until one suffers from a bad one!); (c) prevailing interests of the users, which might contrast with current goals, etc. An obvious consequence is that a "relationship of trust" needs to be built up between the user and agent by allowing the former to understand what the latter knows of the user, and how this information is used. At the same time, a good agent is not only held to make prudent use of any personal information about the user. It is also expected to provide a useful help, and to decide which is the prevailing goal/interest of the user under given conditions.
- *Responsible benevolence*. Unfortunately, such a decision does not depend only upon a scrupulous calculation of the user's utility function. Things are complicated because personal preferences are dynamically re-ordered in consideration of novel information (cf. Cavedon & Sonenberg, 1998), new goals may arise and modify the existing preference relationships, the existing preference order may not correspond to the agent's interests, etc. What is more, sometimes agents are not likely to give up certain goals in favour of others, even when pursuit of the latter would be recommended by the law of utility (their benefit minus their costs is higher than the benefit minus the cost of the former goals). Indeed, a preference order over goals (cf. Castelfranchi & Conte, 1998) often does not allow for a utility-based choice, simply because goals often are not interchangeable!
- *Interactivity and autonomy*. The current direction of development of personal assistants seems to provide an answer to this question in terms of enhanced interactivity. Far from delegating the whole task (*open* delegation in the sense of Castelfranchi and Falcone (1998)), the user should be allowed to keep control over the agent's knowledge and the way this is used. However, this direction of development indicates a potential trade-off between efficiency and interactivity: What is the use of a merely executive assistant that resorts to the user at any step of the task it is responsible for? If applications require a higher level of "responsibility" and "autonomy" of the agent systems, the problems encountered by implementing these systems cannot be solved only by preventing or limiting responsibility and autonomy in task execution. Current applications of expert assistants face an apparent paradox: while demanding higher autonomy of the agent systems, they receive solutions aimed at reducing it! Here, the challenging problem is when, to what extent, and how can any given agent act *on behalf of* another and look over her goals, preferences and interests? What are the boundaries of agents' autonomy, if some autonomy ought to be granted for the sake of efficiency? How to enable agents to take autonomous but sensible and responsible decisions? How should the user's personal knowledge be structured and updated? In other words, what are the mental properties and capacities of a useful autonomous assistant?
- *Social properties and reasons for trust*. This is not to deny the importance of trust relationships in personalised assistance, but to explore the grounds upon which a relationship of trust is constructed. To allow the user to keep control over the agent is only a partial, and not very interesting, solution. People do not entrust others only when they can keep them under strict control. They do so when they have reasons to *believe* that others are trustworthy. What is the content of these beliefs?

First, a trustee is believed, at least to some extent, to be an expert (the CANDO condition) in the task it is delegated (cf. Castelfranchi & Falcone, 1998). Secondly, a trustee is believed to adopt the goal of the trusting agent (the WILL condition). Thirdly, it is believed to commit to the delegated task and to keep to this commitment (let us call them, respectively, the INTEND and the FULFIL conditions).

These beliefs are involved in entrusting someone with a given task. However, the more open the delegation (as seems to be the case with personal assistance), the wider the trust: the trusting agent will resort to open delegation (entrust the agent with a goal or (sub-)plan rather than with a specific action, and let it decide how to achieve it) only when she trusts the delegated agent's "good-will" (Boniver-Tuomela, in preparation), i.e., its intention not to harm the trusting agent's goals, including those not directly involved in the task in question. In the case of automated personal assistants, good-will seems to include (a) preposition, namely the agent's capacity to check its plan and predict possible errors and interference with other goals of the user, and more generally, to evaluate its autonomy with regard to the user's goals, and (b) interactive *meta*-planning, turning to the user *when* the agent believes it cannot be autonomous in solving conflicts and overcoming interference between the user's goals and interests. A personal assistant (which is designed to be benevolent) is trustworthy when it is endowed with the capacity to realistically evaluate its power (including knowledge and a decision-making capacity), to achieve or maintain the user's goals, and the capacity to find ways (including turning to the user) in which to compensate for its possible incapacities.

3.1.2 Expert controllers

Analogous considerations *a fortiori* apply to situations in which the agent controls the user (for example, the way in which she executes her tasks) and influences her (for example, drawing her attention to given information). In addition to trustworthiness, this type of application implies quite complex social capacities.

In controlling the user's task performance, nontrivial social reasoning capacities are essential for *detecting* the user's (possible) deviations from a standard task-execution (which may necessitate inferring the user's next moves). Often, errors should not only be detected, but also *prevented*, especially when dangerous, irreparable or costly mistakes are predictable. Finally, social competencies and decisions are involved in *communicating* the result: *What* should be communicated? Should the agent warn only against possible mistakes, and simply register effective errors? Should it remind about past errors, or only the most frequent, or the most severe ones? *How* should it communicate, by simply conveying information or expressing related emotions? And how should it convey information; by predicting (bad) consequences, reminding about correct performance, stressing conditions for action apparently ignored by the user, etc.? *When* should it communicate, how preventative must a warning be? These are only some of the questions which arise in performance control. Answering them implies some investigation into the issue of social influence, i.e., in the capacity to control and modify the mental states of others. To implement adequate solutions, however simplified, implies that the system is not only provided with knowledge about the user's mental states, but also with the capacity to use this knowledge in order to act upon the user's mind.

3.2 Multiagent applications

In multiagent domains, most problems arise as a consequence of and in order to manage the complexity and reduce the costs of distributed solutions.

3.2.1 Problems

The expansion of applications for software agents depends at least in part upon the impact of these applications on the management of complexity in multiagent domains and tasks.

- *Limits of negotiation*. In competitive contexts (e.g., electronic commerce), the process of

negotiation (even in terms of time budgets) is found to be costly, and elementary forms of organisations are explored. To enhance efficiency and reduce the costs of negotiation, third generation software agents are often endowed with mechanisms for coalition formation (cf. Guttman et al., 1998). Overall, there seems to be a trade-off between decentralisation and efficiency. On the one hand, reliance on decentralised solutions (not only in various multiagent domains, but also in structured organisational environments; cf. O'Brien and Wiegand (1998)) is on the increase. On the other hand, the early DAI emphasis on *negotiation* has been reduced. This is due to two main features of current software agents: (a) (semi-)*autonomous* agents with their local interests protract and complicate the negotiation process; (b) *personalised* agents (cf. Maes, 1994; Guttman et al., 1988) render the negotiation process even more complex at the expense of global efficiency. How may we combine autonomous and personalised benevolence with efficiency? Paradoxically, the more horizontal and distributed the structure of the systems in which software agents are introduced, the less one can rely upon the mere negotiation as a rational process to achieve coordination, although negotiation protocols have been shown to allow self-interested agents to cooperate on tasks (Zlotkin & Rosenschein, 1994; Shehory & Kraus, 1995; Ketchpel, 1995).

- *Complex organisational management*. In cooperative contexts, horizontal structures have been found to be more apt for current dislocated organisations. As is emphasised by some authors (e.g., O'Brien & Wiegand, 1998), organisations are increasingly becoming distributed. At the same time, horizontal organisations raise problems of complex task allocation, control management, etc. The shift from hierarchical to horizontal organisations has not reduced the complexity to be managed. Distributed, flat structures give impulse to labour division on the one hand, and control management on the other. Indeed, a trade-off can be perceived between horizontal organisations and delegated control (*Business Week*, 1993). How can we deal with it?
- *Efficiency in teamwork*. In multiagent task execution and teamwork, in particular, the more complex the task structure, the higher the probability of local errors and the greater the impact on the global result (Jennings, 1995; Kaminka & Tambe, 1998). How can we detect, prevent or repair these errors? How can we take advantage of the multiagent potential in execution monitoring?

Below, it will be argued that no significant impact in these domains can be achieved without strengthening the social competence of agent systems.

3.2.2 Solutions proposed

Several solutions have been thought of, if not yet fully implemented, for (some of) the problems raised above: trust and third-party guarantees (cf. Castelfranchi et al., 1998), reputation (cf. the MIT Media Lab's Kasbah, which incorporates a reputation mechanism called the Better Business Bureau; Chavez et al. (1997)), coalition formation (Guttman et al., 1998), control (either delegated or decentralised; cf. the notion of "agency" in O'Brien and Wiegand (1998)). These solutions share some common features:

1. Developments in the field of multiagent systems demand that greater attention be paid to issues of social scientific concern. This in turn requires a cross-fertilisation between social-scientific disciplines and agent systems research and design.
2. Developments in multiagent systems, whether competitive or cooperative, tend to produce analogous problems (costs of decentralisation, limits of negotiation, etc.) and converge on common solutions (coalitions, conventions, monitoring and control, etc).
3. In both contexts, individual social capacities are necessary ingredients for these solutions to be implemented (interaction and communication are essential for coalition formation, etc.).
4. Solutions are also sought and proposed at the *collective* level (coalitions, conventions, reputation); they are not meant to substitute or prevent, but to implement decentralised solutions. Nonetheless, how can a collective solution apply to a non-cooperative context? What is meant, indeed, by a collective solution?

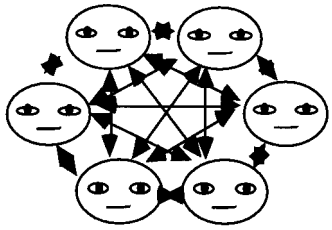

Figure 1 Dependence network in teamwork

3.3 Limited autonomy in multiagent systems

Collective solutions neither imply nor restore benevolence. They are necessary precisely because agents are autonomous and not necessarily benevolent. However, agent autonomy is limited by complex networks of inter-dependence, where agents depend upon one another to achieve their *own* goals. In terms of the dependence theory proposed by some authors (Castelfranchi et al., 1992; Sichman et al., 1994; Conte & Sichman, 1995), an agent x is said to depend upon some other agent y with regard to one of its goals p when (a) x is not autonomous with regard to p: it lacks at least one of the actions or resources necessary to achieve p; while (b) y has the missing action or resource.

Typically, teamwork is an aggregate where each agent depends upon all others to achieve a shared goal (see Figure 1). A market, instead, is a network of dependence, where by definition no agent involved is autonomous with regard to a (subset of its) goal(s), and all may depend upon one another. *Prima facie*, at least, market dependence looks far less structured and less cohesive than teamwork dependence (cf. Figure 2).

Figure 1 depicts quite a complex network, where each agent depends upon all others. Such a network is pre-established, inherent to the team, and is based upon the multiagent task which the team is supposed to accomplish. Figure 2, in contrast, depicts a loose network, where agents' dependence links do not design a common, global dependence. However, several factors enhance the complexity of networks in competitive contexts. The complexity of dependence networks varies across a number of dimensions, including the size of the network, the number and type of connections, etc. (Conte et al., 1998a). This is not the forum for a detailed discussion of these dimensions (but see Conte, 1998). For the purpose of the present discussion, suffice it to say some words about the social goals which arise in competitive environments.

Agents bargaining in the same market are *prima facie* seen as members of a loose, non-integrated dependence network. At most, unilateral (where an agent x depends upon others, but no-one depends upon x) or bilateral dependence (where x depends upon agent y for x's goal p, and y depends upon x for y's goal q) is expected to hold. On the contrary, negative interference and competition for the same partners are expected to abound.

However, even in market-like dependence networks, interesting forms of cohesion take place.

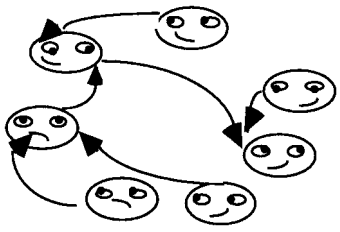

Figure 2 Dependence network in markets

3.3.1 Dependence in markets: Concurrent goals

In a market-like environment, agents share at least one *concurrent* social goal (agents have concurrent goals when each wants to get the same resource for itself; cf. Conte et al., 1991): each agent wants to find partners for exchange. More precisely, each agent depends upon all others for the goal to be a member of (not to be cast out from) the network of potential bargains. Even when no direct partner is currently available, the dynamics of the network and the potentialities of the dependence networks[1] lead one to prefer to stay in the market rather than to be cast out, at least within certain limits (beyond which one may decide to migrate). Consequently, agents in a market each depend upon all others with regard to their concurrent goals to be (perceived as) a potential partner. Reputation in a market shows the same pattern: to find partners in a market, agents need a good reputation. In our terms, it is a concurrent goal, with regard to which agents depend upon one another: can enjoy a good reputation in a given population only if each contributes to preserve it and spread it.

3.3.2 Dependence in markets: Collective goals

The goal to be (perceived as) a potential partner and the goal of a good reputation generate another goal, i.e., that the whole market survives and works. This is an *intrinsically* social, actually intrinsically *collective*, goal: no agent can be autonomous with regard to it. Intrinsically social goals are such that the agents holding them depend upon someone to achieve them. Intrinsically collective goals are such that the agents sharing them each depend upon all others to achieve them. Indeed, the goal that the market be preserved is a collective goal, namely a goal with regard to which (a) each member of the market depends upon all others and (b) the collective result depends upon each member. Agents bargaining in the market share at least one intrinsically collective goal. In this case at least, competition leads to the formation of a collective mental state.

The case of intrinsically collective goals or interests is far from rare.[2] Conventions belong to the same class of phenomena: a convention is effective, it is an advantageous state of affairs, only if the agents whom the convention applies to behave according to it. In other words, to profit from a convention, each agent depends upon all others; but at the same time, the efficacy of the convention depends upon the agents' behavioural regularities.

3.4 Mental requirements for collective solutions

There are different ways of implementing a collective phenomenon, from the simplest to the most complex; from one which does not require social competence, to one which requires social and even collective competence. A collective phenomenon may occur when a set of agents share a given behaviour or feature. In this case, the collective solution is actually pre-established in the agent system, with no need to implement any social competence into the agent. This is the case of conventions implemented as action constraints. The solution is collective, but it implies no social intelligence in the agents in which it is implemented.

However, in such a case the collective solution is achieved at the expense of the agent autonomy. To reconcile autonomy with conventions, agents must be enabled to evaluate, reason upon them and decide whether to accept or violate them (cf. Conte et al., 1998b). This implies that agents have the capacity to form mental representations of conventions and reason upon collectives and collective properties. More generally, for a collective property to apply to autonomous agents, *a collective (i.e., a set of agents) must be represented in the minds of (a subset of) the agents belonging to it, and such a representation must be involved in the process leading to the occurrence and maintenance of that collective* (cf. the notion of joint intentions in BDI agents).

[1]For example, the possibility of *indirect* dependence, where a given agent x, depending on another agent y to achieve its goal p, may turn to depend upon another agent z on which y itself depends for pursuing x's goal p (see Castelfranchi et al., 1992), which are not easily detected by a single agent.

[2]Economists refer to this notion when they speak of "external" goods (cf. Marshall, 1920), or sets of advantageous relationships among individuals in a population.

When and why does collective intelligence have a consequence for individual intelligence? When collective solutions apply to and involve autonomous intelligent agents, they generate further phenomena, such as individual commitment and obligations (cf. Royakkers, 1998, section 5.4.1). For collective phenomena to be implemented into autonomous intelligent agents, these should be endowed with the capacity to derive an individual commitment from a collective one, to acknowledge, accept and comply with collective obligations and derive the corresponding individual obligations. In a few words, to endow autonomous intelligent agents with macro-social competencies implies endowing them with the capacity to reason and decide upon *deontic* representations. Applications of autonomous agents in multiagent domains need to invest in further formal/theoretical work on deontic notions and mechanisms.

4 Artificial social intelligence: What it is (and should not be) and what it should be (and is not)

The notion of social intelligence may suffer from a number of misunderstandings and ambiguities:

1. Social intelligence is sometimes seen as a virtue, a socially desirable inclination; it may be used as equivalent to benevolence and sociability. Consequently, it may be used in opposition to selfishness, and even to autonomy. On the contrary, the importance of this notion will become more apparent when it is seen as a capacity to achieve one's goals, rather than a disposition to be good; as a general aspect of intelligence, rather than a specific trait of personality.
2. Occasionally, authors refer to social intelligence as a systemic phenomenon, an emergent property of multiagent systems. But why conceive of individual agents as generally intelligent and socially deficient? Here, social intelligence is seen as an individual precondition, or prerequisite, for social and collective phenomena to occur.
3. When any individual social intelligence is acknowledged, it often consists of the agents' capacity to interact and communicate with one another. This is a rather limited view. Indeed, social intelligence may be used to communicate and interact with others, as well as to deceive, manipulate and exploit others. Here, social intelligence is meant as a more general capacity than communication and interaction, and refers to the capacities which are needed for interacting and communicating.

To sum up, Artificial Intelligence is here defined as *the study and implementation of agent systems endowed with the capacity to reason upon and modify their own and others' mental states in order to interact with them, and more generally, to act in multiagent contexts in an intelligent and autonomous way*.

5 Conclusions

In this paper, a discussion of current and future developments in the field of agent systems has been presented. As is shown by reports on recent scientific events, the application of agent systems to personalised assistance and multiagent domains is on the increase. It seems reasonable to predict (Rao, 1998) that this will stimulate work on believable agents and knowledge-based assistance. Plausibly, a comparable amount of work will also be invested in the study of social reasoning capacities. Agent systems are requested to accomplish even more complex and demanding social tasks both in interaction and multiagent domains. At the micro-social level, they are delegated tasks which require overhelp and trustworthiness, social control and monitoring. At the macro-social level, they are expected to act as acceptable partners in competition, and responsible members in coalitions. The future challenge of agent systems is to design agents with a capacity to help, compete and cooperate with others in an intelligent, autonomous way.

Artificial Social Intelligence may be seen as a meeting point for several scientific disciplines and fields: the social sciences, the philosophy of mind, AI, and more generally, the science of the artificial, including multiagent systems, believable agents, artificial societies and social simulation, etc. Furthermore, Artificial Social Intelligence is a middle ground on which engineering and formal

theoretical work could meet again. Further investigation in the area of deontic notions and phenomena, as well as in the area of group and collective mental states and actions, is needed. Engineering solutions enabling agent systems to be applied to multiagent domains can be worked out without much theoretical effort. It is doubtful, however, that these solutions allow for implementing agents with a significant level of autonomy. To reconcile agents' autonomy with multiagent efficiency, formal theoretical investigation needs to be integrated with systems implementation. An old idea for a challenging perspective.

References

Boniver-Tuomela, M. *A General Account of Trust*. Technical Report, University of Helsinki, in preparation.

Business Week 1993. "The Horizontal Organisation" 20 December.

Castelfranchi, C and Conte, R, 1998a. "Limits of economic and strategic rationality for agents and MA systems" *Robotics and Autonomous Systems* **24** 127–139.

Castelfranchi, C and Falcone, R, 1998b. "Principles of trust for MAS: Cognitive autonomy, social importance, and quantification" *Proc Third International Conference on Multi-Agent Systems* IEEE, pp. 72–80.

Castelfranchi, C, Cesta, A and Miceli, M, 1992. "Dependence relations in multi-agent systems" in Y Demazeau and E Werner (eds) *Decentralized AI-3* Elsevier.

Castelfranchi, C, Falcone, R, Firozabadi, BS and Tan, YH (eds), 1998. *Pre-proceedings Workshop on "Deception, Fraud, and Trust in Agent Societies"* Minneapolis, St Paul, May 9.

Cavedon, L and Sonenberg, L, 1998. "On social commitment, roles, and preferred goals" *Proc Third International Conference on Multi-Agent Systems* IEEE, pp. 80–88.

Chavez, A, Dreilinger, D, Guttman, R and Maes, P, 1997. "A real-life experiment in creating an agent marketplace" *Proc 2nd International Conference on the Practical Applications of Intelligent Agents and Multi-Agent Technology (PAAM'97)*.

Conte, R, 1998. *Group Dependence* Roma, TR-IP-PSS.

Conte, R and Sichman, J, 1995. "DEPNET: How to benefit from social dependence" *J Mathematical Sociology* **20**(2–3) 161–177.

Conte, R, Miceli, M and Castelfranchi, C, 1991. "Limits and levels of cooperation: Disentangling various forms of prosocial interaction" in Y Demazeau and JP Mueller (eds) *Decentralized AI-2* Elsevier.

Conte, R, Veneziano, V and Castelfranchi, C, 1998a. "The computer simulation of partnership formation" *Computational and Mathematical Organization Theory* **4**(4) 293–315.

Conte, R, Castelfranchi, C and Dignum, F, 1998b. "Autonomous norm acceptance" in J Mueller (ed) *Proc 5th International Workshop on Agent Theories Architectures and Languages* Paris, 4–7 July.

Crabtree, B, 1998. "What chance software agents" *The Knowledge Engineering Review* **13** 131–137.

Etzioni, O, 1993. "Intelligence without robots: A reply to Brooks" *AI Magazine* **14** 7–13.

Goldman, RP and Baral, C, 1998. "Robots, softbots, immobots: The 1997 AAAI Workshop on Theories of Action, Planning and Control" *The Knowledge Engineering Review* **13** 179–185.

Gutmann, RH, Moukas, AG and Maes, P, 1998. "Agent-mediated electronic commerce: A survey" *The Knowledge Engineering Review* **13** 147–161.

Jennings, N, 1995. "Controlling cooperative problem solving in industrial multi-agent system using joint intentions" *Artificial Intelligence* **75** 195–240.

Kaminka, GA and Tambe, M, 1998. "What's wrong with us? Improving robustness through social diagnosis" *Proc National Conference on Artificial Intelligence (AAAI-98)*.

Ketchpel, S, 1995. "Coalition formation among autonomous agents" in C Castelfranchi and JP Mueller (eds) *From Reaction to Cognition* Springer.

Maes, P, 1994. "Agents that reduce work and information overload" *Communications of the ACM* **37** 31–40.

Malsh, T and Schulz-Shaeffer, I, 1996. "Generalized media of interaction and inter-agent coordination" *Pre-Proceedings of the Workshop on Socially Intelligent Agents (SIA)*.

Marshall, A, 1920. *Principles of Economics: An Introductory Volume* MacMillan (8th ed).

O'Brien, PD and Wiegand, ME, 1998. "Agent based process management: Applying intelligent agents to workflow" *The Knowledge Engineering Review* **13** 161–175.

Rao, AS, 1998. "A report on expert assistants at the autonomous agents conference" *The Knowledge Engineering Review* **13** 175–1179.

Royakkers, LMM, 1998. *Extending Deontic Logic for the Formalization of Legal Rules* Kluwer.

Shehory, O and Kraus, S, 1995. "Coalition formation among autonomous agents: Strategies and complexity" in C Castelfranchi and JP Mueller (eds) *From Reaction to Cognition* Springer.

Sichman JS, Conte, R, Castelfranchi, C and Demazeau, Y, 1994. "A social reasoning mechanism based on

Dependence networks'' in AG Cohn (ed) *Proc 11th European Conference on Artificial Intelligence* Wiley, pp. 188–192.

Williams, BC and Nayak, P, 1996. ''Immobile robots: AI in the new millenium'' *AI Magazine* **17**.

Zlotkin, G and Rosenschein, J, 1994. ''Coalition, cryptography, and stability: Mechanisms for coalition formation in task oriented domains'' *Proc National Conference on Artificial Intelligence (AAAI-94)*.