

## Social Attitudes and Personalities in Agents

Cristiano Castelfranchi<sup>1</sup>, Fiorella de Rosi<sup>2</sup> and Rino Falcone<sup>1</sup>

<sup>1</sup> Institute of Psychology, National Research Council,  
Viale Marx 15, I-00137, Rome, Italy  
{cris, falcone}@pscs2.irmkant.rm.cnr.it

<sup>2</sup> Department of Informatics, University of Bari,  
Via Orabona 4, 70126 Bari, Italy  
derosis@gauss.uniba.it

### Abstract

We describe a system (GOLEM) which is aimed at formalising, implementing and experimenting different kinds and levels of social cooperation, represented by different social attitudes or personality traits. First, we examine why personalities in agents are needed and what they are. Second, we propose our definition of the two basic elements of Multi Agent cooperative activity (Delegation and Adoption), by explaining how they are related to the agent's level of autonomy and of cooperativeness. Then, we present how we formalise these levels of Delegation and Adoption in terms of agents' personality traits or attitudes, and we outline how they can be organised in reasonable personalities and interesting interactive situations. Finally, we show how these traits and attitudes are involved in deciding what to do both proactively and in response to the other's social action, and in reasoning about the other's mind.

### Why Do Agents Need Personalities?

Agents endowed with personalities or personality traits, characters, individual attitudes, etc. are spreading around in various domains (Cohen and Levesque, 1990; Lomborg, 1994; Hayes-Roth, 1995; de Rosi et al, 1996; Cesta et al, 1996; Loyall and Bates, 1997). Which are the reasons of this trend? Just curiosity, or is this a necessary development of the "agentification" of AI? There are, in our view, several independent reasons for introducing personalities in Agents. Let's summarise them.

#### a. Social/Cognitive Modelling

One of the major objectives of AI (and ALife) as science is modelling natural intelligence. Since in nature and in society agents have personalities and this seems an important construct in psychology, one might aim at modelling personality in agents (or emotions or cognitive biases) to reproduce relevant features of human interaction.

#### b. Believability and entertainment

Believability has been recognised as one of the most important features for a natural user interaction in entertainment and user-friendly interfaces. It is strongly related to expressing emotions and caricatures (Loyall and Bates, 1997; Walker, 1997) and to reacting in a "typical", or "peculiar" way. Personalities were in fact first introduced in AI to make more "believable" and deceptive some systems like the paranoid PARRY.

#### c. Story and situation understanding

In making the required inferences for understanding a story or a situation, it is necessary not only to know the appropriate scripts and frames and the agents' intentions and

beliefs, but also their personalities. The first quite complete and formal theory of personality was introduced by Carbonell (Carbonell, 1980) for this purpose. As Carbonell says: "Whenever a story includes character development of one of the actors, this development turns out to be useful and often crucial in formulating an understanding of the story." (p.217). This claim is also connected to what is now called "agent modelling": not only in stories but also in real interactions (in human or virtual reality) "knowledge about personality traits is necessary to understand the actions" of the agents.

#### d. Agent Modelling

User stereotypes and profiles proved to be useful in adaptive and cooperative human-machine interaction, to make correct ascriptions and abductions (Rich, 1989). The same is true in multiple agents' interaction. We therefore need defining agents' classes and stereotypes, some of which are personality-based. For example: in user modelling, *student* denotes a role, whereas *aristocratic* or *thrifty* denote personality traits. Among agents, we might have classes like *mediator* or *executive agent* or *information filtering agent*, but also classes like *benevolent* or *self interested*, which correspond to social personality traits or attitudes. All these are interesting reasons for introducing/modelling personalities in the agents. But we believe that there is some more principled reason that holds in the very basic philosophy of agent-based computing: its decentralised character, its open world assumption (Hewitt, 199&), its "experimental" approach.

#### e. Exploring and comparing strategies

One of the most interesting aspects of decentralised and MA systems is that they provide a scenario for experimental exploration of coordination mechanisms, behavioural strategies and organisational structures which could not be designed or predicted by centralized rational planning. This is crucial in the "open system" perspective that characterises the new AI of the '90s (Bobrow, 1991; Hewitt, 1991). Exploring different behavioral, reactive or planning strategies in multiagent systems can be seen as exploring adaptivity, efficiency and coexistence of different "personalities" in agents' models. Personalities had already been introduced in the different kinds of commitment defined by Cohen and Levesque (Cohen and Levesque, 1990) or by Rao and Georgeff (Rao and Georgeff, 1991). No strategy can be defined as a-priori optimal, since the world is open, it changes, it is uncertain and unknown, and since other agents in the world will adopt strategies that

might unexpectedly change the result of our actions. Thus, the new paradigm tends to be in favour of "experiments" and heterogeneity: different solutions to a problem, different reactions to a situation, different ways of reasoning, different priorities in goals, etc are allowed to compete or coexist. Intelligence and efficiency tend to be seen (i) as emergent at the global level rather than being embedded in the individual rules, (ii) as selected post hoc or (iii) as reduced to multiple dimensions and let coexist with less efficient strategies that adapt to changing situations. Heterogeneity is a very good explorative strategy and a very robust adaptive approach. Indeed, different solutions to a problem, different reactions to situations, different ways of reasoning, different priorities in goals, etc. are just "personalities".

#### f. Internal states and behaviour

Agents have "internal states" (Shoham, 1993) which affect their reaction to a given stimulus or their processing of a given input. This seems to be one of the important differences between an "agent" and a piece of program, a software component, a module or a function. Agents, then, react in different ways or give different process results to the same input, depending on their internal state: they have different reactive styles, either stable or transitory. Personality is just a specification of this general property. In agent-based computing, the introduction of personality will therefore be motivated also by the need to introduce different treatments of the same input or different processing reactions, that cannot be decided on the basis of external parameters and tests, or input conditions. These different "computations" are conditional to "internal" parameters which evolve independently of the agent sending the input and are unpredictable. Personalities are only an extreme of this feature: a stable set of (potentially transitory) internal states, acting and reacting modalities, forms of reasoning on the input. When these local and internal states and parameters cannot be reduced in terms of knowledge or ability (which can both be acquired), then they may be seen as "personality traits".

### What is Personality

We will call *personality trait* (Carbonell, 1980) any internal state or processing mechanism of the agent that:

- differentiates it from other agents with which it is interacting or is compared;
- is relatively stable (either built in or inborn or learned, but now quite permanent) and cannot be just adopted or learned from outside on line;
- is mental;
- has to do with motivations, with the way of choosing, of reasoning, of planning and so on.

We agree with Carbonell that personalities are mainly goal based: some of them directly consist in the presence of a typical motivation or in a special importance of a given goal (ex. *sadic*, *glutton*); others can be considered as implicit goals or preferences (see later). However, other personalities are rather based on "cognitive styles": ways of reasoning, attending, memorising, etc.

### Personality Traits and Attitudes

That personality traits are stable does not mean that they are continuously relevant or active: if *x* is a *glutton*, when he is working this can be irrelevant. Some personality traits are conditional to a given circumstance: they are just temporary *attitudes*. An attitude is characterized by tests/conditions specifying the circumstance for its activation. An agent can assume an attitude or another (relatively to the same problem) depending on circumstances or partners. Therefore, we distinguish in GOLEM between two constituents of personalities: *traits* and *attitudes*. An agent can change or decide about its attitude towards a given event, request, or agent, while it cannot change or decide about its personality traits: these are not subject to contextual changes or decisions. In presenting GOLEM personalities, we will first introduce some personality traits, which are not tuned to situations or interactions. Later, we will show how these traits could become more flexible social attitudes, by giving the agent the possibility of adopting it or not, depending (for example) on the partner's personality.

In short: *a personality is a coherent believable, stable, and typical cluster of traits and attitudes that are reflected in the agent's behaviour.*

### Personality and Emotions

Emotional states are among those internal states that shape an agent's cognitive process and reaction; they can also characterise the agent. Emotion-based personalities can be defined, like *shameful*, *fearful*, *pityful* and so on: these personalities are characterised by the agent's propensity for a given emotional reaction. However, emotions and personalities should not be mixed up with one the other, like it risks to happen in the "believable agent" domain. This is due to the fact that, in that domain, personalities are introduced just for the sake of "believability", and believability for sure requires emotional reactions (Elliott, 1994; Hayes-Roth, 1995; Picard, 1996; Loyall and Bates, 1997). In our view:

- emotions do not necessarily imply personalities, since there might be emotional behaviours that are shared by the whole population of agents and do not characterise particular agents or individuals;
- personalities are not necessarily related to emotions: they might be just based on (i) cognitive properties or styles, like a "fantasyful" agent, or a "fanatic" one, (ii) preferences and goals, (iii) interactive strategies (ex. Tit-for-Tat agents; or cheaters, etc.).

Of course, it is true that these cognitive styles, and in particular preferences and goals, can make a given type of agent exceptionally liable to some emotions. However, these emotions are not the basis for constructing and characterising that agent, though being useful to recognise it. In addition, emotions are not necessary: agents might be free from emotions while having personalities.

### Delegation and Adoption Levels

Delegation and Adoption are the two basic ingredients of

any collaboration and organization. A huge majority of DAI and MA are based on the idea that *cooperation works through the allocation of some task of a given agent to another agent*, via some "request" (offer, proposal, announcement, etc.) meeting some "commitment" (bid, help, contract, adoption, etc.). In (Castelfranchi and Falcone, 1997), an analytic theory of delegation and adoption was developed to contribute to understanding and clarifying the cooperative paradigm. Informally:

- *in delegation, an agent A needs an action of another agent B and includes it in its own plan.* In other words, *A is trying to achieve some of its goals through B's actions; thus A has the goal that B performs a given action.* A is constructing a MA plan and B has a share in this plan.
- *in adoption, an agent B has a goal since and until it is the goal of another agent A, i.e. B has the goal of performing an action since this action is included in A's plan.* So, also in this case B plays a part in this plan.

Both delegation and adoption may be unilateral: B may ignore A's delegation while A may ignore B's adoption. In both cases, A and B are, in fact, performing a MA plan.

One can distinguish among at least the following types of delegation:

- *pure executive delegation Vs open delegation;*
- *strict delegation Vs weak delegation;*
- *delegation Vs non delegation of the control over the action;*
- *domain task delegation Vs planning task delegation (meta-actions)*
- *delegation to perform Vs delegation to delegate.*

For an accurate analysis of these dimensions of delegation, see (Castelfranchi and Falcone, 1997). Let us here consider the dimensions which characterize the *autonomy* of the delegated agent (B) from the delegating one (A):

- level of delegation 'openness',
- level of control of actions given up or delegated,
- level of decision left to B,
- level of dependence of B on A, as for the resources necessary for the task,

The object of delegation can be specified minimally (*open delegation*), completely (*close delegation*) or at any intermediate level. We wish to stress that *open delegation* is not only due to A's preference, practical ignorance or limited ability. Of course, when A is delegating a task to B, he is always *depending on B* for that task: he needs B's action for some of his goals. However, open delegation is also due to A's ignorance about the world and its dynamics: *fully specifying a task is often impossible or not convenient.* Open delegation is one of the bases of the *flexibility* of distributed and MA plans. In analogy with delegation, several dimensions of adoption can be characterized. In particular:

- *Literal help:* B adopts exactly what was delegated by A (elementary or complex action, etc.).
- *Overhelp:* B goes beyond what was delegated by A, without changing A's plan.
- *Critical help:* B satisfies the relevant results of the

requested plan/action, but modifies it.

- *Overcritical help:* B realizes an Overhelp by, at the same time, modifying or changing the plan/action.
- *Hyper-critical help:* B adopts goals or interests of A that A itself did not consider; by doing so, B does not perform the action/plan, nor satisfies the results that were delegated.

It is then possible to define the *level of collaboration* of the adopting agent: there are agents that help other agents by just doing what they were literally requested to do; there are agents that have initiative, have care of others' interests: they use their knowledge and intelligence to correct others' plans and requests that might be incomplete, wrong or self-defeating.

## GOLEM's Architecture and Language

GOLEM is a factory of proactive agents which cooperate, in a 'play', to transforming a world. Each agent has:

- its *goal* (state of the world that it desires to achieve),
- its *knowledge* of the present world-state and of how the world can be transformed (world-actions that can be performed in each world-state)
- its *know-how* about world-actions.

The agent has also a *personality*, as a set of traits and attitudes which affect its behaviour in the different phases of the 'play'. All agents are entitled to perform *world-transformation* or *world-control actions*, as well as *communicative actions*: they communicate by exchanging speech acts and by looking at the results of actions performed by others [Ric97]. In the present prototype, we limit the play to two agents. However, considerations and simulation methods can be easily extended to the multi-agent case. The play begins with the definition of the initial state of the world and of the agent which moves first. The two agents alternate in a regimen of 'turn taking'. At each turn, the playing agent decides, after a reasoning process, which (communicative or world) action to perform. In this process, it applies various forms of reasoning: goal-directed inference, cognitive diagnosis, plan evaluation, goal recognition, ATMS-based knowledge revision. The knowledge base employed includes a description of own mental state and a default image of the other agent's mental state. Let us now describe the main components of an agent's mental state:

### a. atomic beliefs and goals

a.1. BEL Ai p, with p = positive or negative literal which represents one of the following attitudes:

- *know-how:* (CanDo Aj a); the agent Aj can execute a world-action a;
- *intention:* (IntToDo Aj a);

Aj intends to perform a during the present turn.

a.2. GOAL Ai g, with g = world-state or literal about another agent's intentional state.

Our agents do not hold only the problem of deciding by themselves, but have, as well, desires about other agents' intentions: they consequently may have the purpose of influencing other agents' behaviour so as to satisfy these desires.

### b. rules

While atomic beliefs about own and other agents' know-how are part of an agent description, as well as its goal and its domain knowledge, the agent intentions and its desires about other agents intentions are the logical consequences of these 'basic beliefs': relationships among these elements are represented in three sets of rules, that we call *strategies*. Different strategies are applied when reasoning on whether to delegate some action (*delegation strategies*), whether to provide an explicitly or implicitly requested help (*help strategies*), or how to react to another agent's helping decision (*reaction strategies*). As reasoning by an agent is aimed at deciding 'what to do next', we do not need representing time in our language. This means that, in GOLEM, agents cannot intend to perform a specific action in a specific time instant: they can only decide whether or not to do that action *at their turn*. Desires and intentions established as logical consequences of basic beliefs and strategies are transformed into communicative actions or into world transformation or control actions by three sets of *commitment rules*:

b.1. *in the delegation process*, intention to perform a world-action *a* is transformed into execution of *a*, provided that *a* is 'performable'. The desire to influence another agent's intention to perform *a* is translated into a 'directive' or an 'assertive' communicative act which depends on what is presumed to be the present intentional state of that agent. In case of doubt about the other agent's intentions, an interrogative act is made. The belief that no intentional state in favour of *a* can be achieved sets the agent in a 'waiting' state.

b.2 *in the helping process*, own intention to perform the action requested by another agent is transformed into a 'declarative' and a 'commissive' communicative act, followed by the immediate execution of the domain-action: in the negative intention case, only a declarative act of 'refusal' is made.

b.3. *in the reaction phase*, the helper's manifestation of disagreement about performing the requested action may produce, in the delegating agent, a 'commissive' act, the research of an alternative plan or the decision to set itself in a 'waiting' state; spontaneous offers of help are evaluated, to be accepted or rejected by a 'declarative' act.

## Personalities in GOLEM

A personality trait or attitude in GOLEM is a set of stable preferences, represented as a consistent block of delegation, adoption or reaction strategies.

### Delegation Attitudes

As we mentioned in Section 5, reasoning about delegation ends up with: (a) the agent's intention to do a specific action by itself, (b) the desire to induce that intention on the other agent by delegating that action or (c) the decision to renounce to that action, by 'waiting'. Personality traits establish a "preference rate" among the three alternatives, in the form of strategies to decide which one to select. For example:

- a *lazy* agent always delegates tasks if there is another agent which is able to take care of them; it acts by itself only when there is no alternative;
- a *hanger-on* will never act by itself (Cesta et al, 1996);
- a *delegating-if-needed* asks for help only if it is not able to do the task by itself;
- a *never-delegating* considers that tasks should only be achieved if it can perform them.

As we said, these personality traits are represented as a consistent block of strategies and commitment rules. For example, a *lazy* delegating agent *A* has the following strategies:

If *A* has the goal that a given action *a* is performed for its goal *g*, and it believes that another agent *B* can do *a*, then *A* has the goal that *B* intends to do *a*.

If *A* has the goal that a given action *a* is performed for its goal *g*, and it believes that no other agent *B* can do *a*, and *A* cannot do *a*, then *A* believes that *g* is not currently achievable.

If *A* has the goal that a given action *a* is performed for its goal *g*, and it believes that no other agent *B* can do *a*, and *A* can do *a*, then *A* intends to do *a*.

### Helping Attitudes

Reasoning about adoption may end up with two alternatives: to help or to refuse helping. At least two personality traits influence establishing whether and how to help:

a. *level of propensity* towards helping: various factors contribute to deciding whether to help: (i) own know-how, (ii) presumed know-how of the delegating agent and (iii) consistency of the required action with own goals. Personality traits establish priorities for these factors:

- a *hyper-cooperative* always helps if it can;
  - a *benevolent* first checks that the request does not conflict with its goal;
  - a *supplier* first checks that the other agent could not do the action by itself;
  - a *selfish* helps only when the requested action achieves its own goals (Cesta et al, 1996).
  - a *non-helper* does never help, by principle.
- b. *level of engagement* in helping: the helper may interpret the received delegation according to how much it really wants to meet the delegating agent's desires.
- a *literal helper* restricts itself to considering the requested action;
  - an *overhelper* hypothesizes the delegating agent's higher order goal, and helps accordingly;
  - a *subhelper* performs only a part of the requested plan;
  - a *critical helper* modifies the delegated plan.

Other attitudes further diversify the behaviour of the helping agent:

- *control of conflicts* between the requested action and its own goals: an action can immediately bring to a state which is in conflict with the helper's goal state (in a situation of 'surface-conflict') or it can be part of a delegating agent's plan which, in the long term, will produce a conflict (in a

situation of 'deep-conflict'). A *deep-conflict-checker* will check that no such conflicts are created by the requested action, by making a goal-recognition on the delegating agent's mental state. A *surface-conflict-checker* will only examine the immediate consequences of the requested action.

- *control of the delegating agent's know-how*: this, again, can be restricted to examining whether that agent would be able to perform the requested action (in a *surface-knowhow-checker*) or can investigate whether alternative plans exist, which bring to the delegating agent's presumed goal and that this agent would be able to perform by itself (in a *deep-knowhow-checker*).

## Reaction Attitudes

Agents are governed, in their reactions to offer or refusal of help, by the same attitudes which rule out their delegation decisions:

- a *lazy* agent will perform the world-action that it attempted to delegate, if the helper rejected this request;
- a *delegating-if-needed* will, on the contrary, reject offers of actions that it can do by itself;
- a *never-delegating* will reject any offer of help, ...and so on.

## Some Consistent Personality Traits

Agents' profiles are defined as a combination of delegation, helping and reaction traits and attitudes: this corresponds to the well known stereotype-based approach to modeling, in which multiple inheritance is exploited to produce a multi-faceted representation of a user or an agent (Rich, 1989). However, not all combinations of attitudes and traits are logically consistent. Some of them are implausible: for instance, a *hanger-on* is necessarily also a *non-helper*. Others are unlikely to happen; for instance: one might imagine a *generous* personality, that is very open to help the others but does not want to disturb them as far as it can do it by itself. However, in a more reasonable combination, a *delegating-if-needed* is also a *supplier*, which delegates only if it cannot do, and adopts only when the others cannot do and the requested plan does not conflict with its own goal. As one of the objectives of GOLEM is to evaluate the rationality and the believability of coexisting social attitudes, when different agents are created, they are given personality traits which produce interesting to investigate social interactions. For example: if two *never-delegating&selfish* agents meet, no delegation and no help will be made; if two *hanger-on&selfish* agents meet, no cooperation attempt will succeed.

## Agent modelling: knowing the other

In order to be really helpful and, in general, to better interact with other agents, an agent needs some representation of the other agent's goals, intentions and know-how. This can be based on personal acquaintance, on memory of interactions (Lomborg, 1994), on reputation or on self-presentation. It can be inherited by the class to which the agent is known or

presumed to belong or can, finally, be abducted from the other agent's practical and communicative behaviour (de Rosi et al, 1996).

In the present version of GOLEM, our agents introduce themselves at the beginning of the play, by describing their world-goals, their know-how and their personality: this description can be partial (some features are omitted) or fuzzy (others are described in abstract terms) but is always consistent.

An example: in a blocks world, an agent can describe itself as a *lazy* person who is able to handle blocks of small dimensions and would like to build a twin tower" (that is, a domain state in which a small-blocks-tower and a big-blocks-tower coexist). More generically, it might introduce itself as "someone who tends to delegate tasks and likes complex structures". In this second formulation, it is not clear whether the agent is *lazy* or just *hanger-on* and whether its domain-goal is a twin-tower or some state in which several structures coexist. The agent can also omit personality traits from this description. As a sincere-assertion assumption underlays the system, the other agent acquires the described features to build up a first image of the other agent, and eventually updates this image during interaction by applying several forms of abductive reasoning.

## Abducting the Other Agent's Personality

Our agents can apply their reasoning capability to their general knowledge about personality traits, for abducting the personality of their partners from their behaviour. Let us give some examples.

### a. Abduction about delegation personalities

If the other agent, that I presume to be able to do the required action, after my refusal of help does not do that action by itself,

then I may presume that it is a *hanger-on*;

If the other agent did a given action a1 for a plan p1, and delegates another action a2, and

I presume that it is not able to do a2,

then I may presume that it is either a *delegating-if-needed* or a *lazy*.

### b. Abduction about helping personalities

If the other agent accepts to perform a delegated action and I know that there is a surface-conflict between that action and what I presume (or know) to be its domain-goal,

then I may presume that it is a *hyper-cooperative*;

If the other agent refuses to adopt my delegation though being able to do the action, and I know that it believes that I'm not able to do it and that no conflicts exist between this action and its goals,

then I may presume that it is a *supplier*

## Abducting the Other Agent's Mental State from its Personality

Knowledge of other agents' personality plays a relevant role in abducting their capabilities or plans. Some examples:

If the other agent is a *hyper-cooperative* and refuses my

delegation,  
**then** I can presume that it is not able to do the action  
**If** the other agent is *benevolent* and I presume (or know) that it is able to do the action, and it refuses to help me,  
**then** I can presume that the action I try to delegate produces some conflict with its goal.  
**If** the other agent is a *supplier*, refuses my delegation, and I presume (or know) that it is able to do the action, and I know that this action does not conflict with its goals,  
**then** I can presume that it believes that I'm able to do the action by myself.  
**If** the other agent is a *delegating-if-needed* and delegates an action,  
**then** I may presume that it is not able to do that action.

### Personality-based abduction

Attachment of a plausibility degree to alternative explanations, during the abduction process, can be influenced by the personality of the reasoning agent. In absence of other information, selection of the 'most plausible' hypothesis, in plan recognition or cognitive diagnosis, is guided by attitudes which originate from the relationship between the reasoning agent and the agent it is reasoning about. Two examples:

**If** you requested me to perform the action and I'm *benevolent* and I've no information about your goals and, from my plan recognition process, I can make two hypotheses about your final goal-state, one of which is in conflict with my own goals, whereas the other is not and I'm *suspicious*  
**then** I will select the hypothesis of conflict and will not help you.  
**If** you requested me to perform the action and I'm a *supplier* and I've no information about your know-how and personality and, from my cognitive diagnosis process, I can make two hypotheses about the reasons behind your delegation, one of which being that you could do the requested action but are *lazy*, the other that you cannot do it and are *delegating-if-needed* and I'm *trustful*  
**then** I will select the second hypothesis and will help you.  
 These examples show that new personality attitudes affect this aspect of reasoning: being *suspicious* or *trustful* may be a permanent trait, but also an attitude induced by the other agent's previous behaviour.

### Future Developments

We will provide experimental evidence of statements made in this paper by exhibiting results of simulations with different personality combinations, and will evaluate the believability of these personalities in interactive exchanges. New social personality traits need also to be included in GOLEM. For example, personalities based on typical goals (like "agents whose goal is to build some kind of tower", in the blocks world) or on propensity to deceive. We are particularly interested to investigate the behaviour of agents

that tend to propose an "exchange", either spontaneously or in response to some delegation. We are working also to a systematisation of traits and attitudes in an inheritance hierarchy, in reasonable complex personalities and in M-A situations relevant for both the theory and the application of cooperative systems.

### References

- Bobrow, D.1991. Dimensions of Interaction. *AI Magazine*, 12, 3, 64-80.
- Carbonell,J.1980. Towards a process model of human personality traits. *Artificial Intelligence*, 15.
- Castelfranchi,C. and Falcone,R. 1997. Delegation Conflicts. *Proceedings of MAAMAW '97*.
- Cesta, A., Miceli,M. and Rizzo,P. 1996. Effects of different interaction attitudes on a multi-agent system performance. In W. Van de Welde & J W Perram (Eds.), *Agents breaking away*. Springer-Verlag.
- Cohen,P.H. and Levesque,H.J. 1990. Rational interaction as the basis for communication. In P R Cohen, J Morgan and M E Pollack (Eds): *Intentions in Communication*. The MIT Press.
- de Rosi, F., Grasso, F., Castelfranchi, C. and Poggi, I. 1996. Modeling conflict resolution dialogs between believable agents. *ECAI Workshop on Conflicts in AI*.
- Elliott, C. 1994. Research problems in the use of a shallow Artificial Intelligence model of personality and emotions. *Proceedings of the 12<sup>th</sup> AAAI*.
- Hayes-Roth, B. 1995. Agents on stage: advancing the state of the art of AI. *Proceedings of IJCAI95*
- Hewitt, C. 1991. Open information systems semantics for distributed artificial intelligence. *Artificial Intelligence* 47, 79-116.
- Lomborg, B. 1994. Game Theory vs. Multiple Agents: The Iterated Prisoner's Dilemma. In C Castelfranchi and E Werner (Eds.) *Artificial Social Systems - MAAMAW'92*, Springer-Verlag LNAI 830, 69-93.
- Loyall, A.B. and Bates, J.1997. Personality-Rich Believable Agents That Use Language. In *Proceedings of Autonomous Agents 97*, Marina Del Rey, Cal.,106-13.
- Picard, R. 1996. Does HAL cry digital tears? Emotion and computers.
- Rao, A S. and Georgeff, M.P. 1991. Modeling rational agents within a BDI-architecture. In *Principles of Knowledge Representation and Reasoning*.
- Rich, C. and Sidner, C.L. 1997. COLLAGEN: When Agents Collaborate with People. In *Proceedings of Autonomous Agents 97*, Marina Del Rey, Cal., 284-91.
- Rich, E. 1989. Stereotypes and user modeling. In: *User Models in Dialog Systems*. A Kobsa and W Wahlster (Eds), Springer-Verlag.
- Shoham, J. 1993. Agent-oriented programming. *Artificial Intelligence*, 60.
- Walker, M.A., Cahn J.E. and Whittaker S.J. 1997. Improvising Linguistic Style: Social and Affective Bases of Agent Personality. In *Proceedings of Autonomous Agents 97*, Marina Del Rey, Cal., 96-105.