

# The Human in the Loop of a Delegated Agent: The Theory of Adjustable Social Autonomy

Rino Falcone and Cristiano Castelfranchi

**Abstract**—In this paper, we refer to social autonomy in a collaborative relationship among agents based on delegation and help. We address the problem of adjustable autonomy, i.e., we discuss when modifying the assigned/received delegation entails a corresponding enlargement or restriction of autonomy and at which level. We stress in particular the role played in autonomy by: 1) the degree of “openness” of delegation (*execution autonomy*), 2) the allowed initiative in (re)starting negotiation (*meta-autonomy*), 3) the degree and kind of control (*feedback+intervention*), and 4) the strength of delegation with respect to interaction. We show how the adjustability of delegation and autonomy is actually “*bilateral*,” because not only the user (delegator, trustor, client) can adjust the autonomy of the agent (deleegee, trustee, contractor), but the agent can also have (cooperative) reasons for and the ability to change the received delegation and modify its own autonomy in it. Adjustment is also “*bidirectional*” (from more autonomy to less autonomy, or vice versa), and *multidimensional*. Finally, we analyze some reasons for modifying the assigned autonomy and show how the adjustment of autonomy depends—on the delegator’s side—on a <<crisis of trust>>; vice versa, the deleegee’s adjustment of its own autonomy depends on some disagreement about the trust received from the delegator, and, in particular, either a higher or lower confidence in itself or in external circumstances. Some preliminary hints about necessary protocols for adjusting the interaction with agents are provided. This work is neither directly prescriptive nor simply descriptive. It is aimed at providing a theoretical framework, i.e., the conceptual instruments necessary for analyzing and understanding interaction with autonomous entities.

**Index Terms**—Adjustability, autonomy, control, delegation.

## I. INTRODUCTION

**I**N THIS PAPER we are going to analyze the complex scenario in which a cognitive agent (an agent with its own beliefs and goals) has the necessity to decide if and how to delegate or adopt a task for another agent in a given context, how much autonomy is necessary for a given task, how *trust* and *control* play a relevant role in this decision, and how important their relationships and reciprocal influences are.

*Autonomy* is very useful in cooperation (why should we have an intelligent collaborator without exploiting its intelligence?) and even necessary in several cases (situatedness, different competence, local information and reactivity, decentralization, etc.).

Manuscript received December 21, 2000; revised June 22, 2001. This work was supported in part by the ALFEBITE European Project Contract Number IST-1999-10298 and in part by the TICCA Project (joint research venture between the Italian National Research Council (CNR) and Provincia Autonoma di Trento).

R. Falcone is with the Division of Artificial Intelligence, Cognitive, and Interaction Modeling, Institute of Psychology, Italian National Research Council (CNR), Rome, Italy (e-mail: falcone@ip.rm.cnr.it).

C. Castelfranchi is with the Department of Communication Sciences, University of Siena, Siena, Italy (e-mail: castel@ip.rm.cnr.it).

Publisher Item Identifier S 1083-4427(01)08259-5.

but it is also risky because of misunderstandings, disagreements and conflicts, mistakes, private utility, etc. A very good solution to this conflict is maintaining a high degree of interactivity *during* the collaboration, by providing *both* the man/user/client and the machine/deleegee/contractor the possibility of having initiative in interaction and help (*mixed initiative* [1], [2]) and of *adjusting* [3] the kind/level of delegation and help, and the degree of autonomy run time. We will analyze a specific view of autonomy which is strictly based on the notions of delegation and adoption [4]. In fact, in several situations the multi-agent plan, the cooperation between the delegating agent (*delegator*) and the delegated one (*deleegee*) requires a strict collaboration, and a control flow between the partners, in order to either maintain the delegator’s trust [5] or avoid breakdowns, failures, damages, and unsatisfactory solutions.

Software and autonomous agents will not be useful only for relieving human agents from boring and repetitive tasks; they will be mainly useful for situations where delegation and autonomy are necessary (“*strong dependence*,” [6]) because the user/client/delegator does not have the local, decentralized and updated knowledge, or the expertise, or the just-in-time reactivity, or some physical skill that requires some local control-loop. Thus autonomy and initiative are not simply optional features for *agents*; they are necessary requirements, and obligatory directions of study. However, *control cannot be completely lost and delegation cannot be complete*, not only for reasons of confidence and trust, but for reasons of distribution of goals, of knowledge, of competence, and for an effective collaboration. In this sense the possibility to control and adjust the autonomy of the agents is becoming a growing and interesting research field [3]. Our claim is that: *in designing how to adjust the level of autonomy and how to arrive to a dynamic level of control, it is necessary to have an explicit theory of delegation (and trust), which specifies different dimensions and levels of delegation, and relates the latter to the notion and the levels of autonomy.*

Thus, we propose our plan-based analysis [7] of levels of delegation and levels of help, and discuss a related notion of autonomy. In several cases of collaboration among agents an *open delegation* is required, that is the delegation <<to bring it about that...>>. The agent is supposed to use its knowledge, intelligence, ability, and to exert some degree of discretion.

Moreover, since the knowledge of the client about the domain and its helping agents is limited (possibly both incomplete and incorrect), some *delegated task* (the requested or the elicited behavior) might not be so useful for the client itself while perhaps the contractor is able to provide greater help with its own knowledge and ability, going beyond the *literally* delegated task. We will call *extension of help* or *critical help* this kind of help.

However, of course, there is a trade-off between pros and cons both in *open delegation* and in *extended (critical)-help*: the more intelligent and autonomous the delegee (able to solve problems, to choose between alternatives, to think rationally, and to plan), the less *passively obedient* it is.<sup>1</sup> So, possible conflicts arise between a client and its contractor; conflicts which are due either to the intelligence and the initiative of the contractor or to an inappropriate delegation by the client. We are interested here only in the conflicts originating from the agent's willingness to collaborate and to help the other in a better and more efficient way: a kind of *collaborative conflicts*. We do not consider the contractor's *selfish reasons* for modifying delegation (because the nature of the conflict, negotiation, etc. would be different).<sup>2</sup>

It is worth specifying that this work is neither directly prescriptive—providing guidelines for building multi-agent systems (MAS) or designing human-machine interaction (HMI)—nor simply descriptive. It is aimed at providing a theoretical framework, i.e., the conceptual instruments necessary for analyzing and understanding interaction with autonomous entities. We assume that this framework is useful also for principled engineering, i.e., for getting systems designed not only on the basis of empirical data and practical experience, but also on the basis of a more complete view and typology, and of some prediction and explanation. We also suggest some criteria about when and why an adjustment of the autonomy of an agent should be made, and preliminary hints about necessary protocols for adjusting the interaction with agents.

In Section II-A, we present our delegation/adoption theory and the various dimensions of delegation and help, in particular the interaction and specification dimensions. This (quite long) analysis is very relevant because it introduces a very large and multidimensional spectrum of kinds and levels of task allocation and of helping another agent to achieve its own goals; it is also shown how these possibilities are, in fact, characterized by the mental components of the agents involved in the interaction.

In Sections IIIA-E, we analyze how both delegator and delegee can adjust a previous delegation/adoption action. We see what adjustments are possible and which of them are really interesting. We also analyze the channels and the ways of interaction.

In Sections IV-A and B, the relationships between autonomy and delegation/adoption are shown.

In Sections V-A–D, we consider classes of reasons and principles on which adjustments are based.

Finally, in Section VI-A, we describe how those reasons affect, in fact, the trust of either the delegator or the delegee.

## II. NOTION OF AUTONOMY IN COLLABORATION

For the purpose of this paper we use a practical and not very general notion of autonomy.<sup>3</sup> In particular, we refer to the *so-*

<sup>1</sup>Obviously, a very autonomous but stupid agent is even worse.

<sup>2</sup>In this paper on autonomy, there is at least one reason that should be taken into account: when the contractor adjusts the delegation for having more autonomy.

<sup>3</sup>We do not consider here some important aspects of autonomy (that could be adjusted) like the agent's independence or self-sufficiency. For an analytical discussion on the notion of *autonomy* in agents and for a more principled definition, see [8]–[10].

*cial* autonomy in a *collaborative* relationship among agents. We distinguish between

- a *metalevel autonomy* that denotes how much the agent is able and in a condition of negotiating about the delegation or of changing it (to this regard, a slave, for example, is not autonomous: he cannot negotiate or refuse);
- a *realization autonomy* which means that the agent has some discretion in finding a solution to an assigned problem, or a plan for an assigned goal.

Both are forms of goal-autonomy, the former at the higher level, the latter at the subgoals (instrumental) level.

For a definition of different kinds of autonomy, including some of the dimensions we consider, see also [11].

The lower is the client's control (monitoring or intervention) on the execution, the more autonomous is the contractor. In this context then, *autonomy means the possibility of displaying or providing an unexpected behavior (including refusal) that departs from the requested (agreed upon or not) behavior. The autonomous agent can be either entitled to or not to perform such an unexpected behavior.*<sup>4</sup>

### A. Delegation/Adoption Theory

1) *Formal Constructs*: Several formal constructs are needed in the following. Let  $\text{Act} = \{\alpha_1, \dots, \alpha_n\}$  be a set of *actions*, and  $\text{Agt} = \{\text{Ag}_1, \dots, \text{Ag}_m\}$  a set of *agents*. The *general plan library* is  $\prod = \prod^a \cup \prod^d$  where  $\prod^a$  is the abstraction hierarchy rule set and  $\prod^d$  is the decomposition hierarchy rule set. An action  $\alpha' \in \text{Act}$  is called *elementary action* in  $\prod$  if there is no rule  $r$  in  $\prod$  such that  $\alpha'$  is the left part of  $r$ . We will call *basic actions* ( $B\text{Act}$ ) the set of elementary actions in  $\prod$  and *complex actions* ( $C\text{Act}$ ) the remaining actions in  $\text{Act}$ .

Given  $\alpha_1, \alpha_2$ , and  $\prod^d$  we introduce the  $\text{Dom-c}(\alpha_1 \alpha_2)$  operator to say that  $\alpha_1$  *dominates*  $\alpha_2$  (or  $\alpha_2$  is *dominated* by  $\alpha_1$ )

<sup>4</sup>In this paper, for the sake of brevity, we do not discuss in detail another very important distinction between

- being *practically in condition* of doing something (refusing, negotiating, changing, and doing something else), i.e., what we would like to call  $\ll \text{practical possibility} \gg$ ;
- being *deontically in a condition* of doing something, i.e., to be entitled, permitted in the strong sense, i.e., the  $\ll \text{deontic possibility} \gg$ .

An agent can have the former without the latter, or vice versa (see [10]). In fact, there are two kinds of lack of power (hence, of dependence and autonomy): one based on practical conditions, the other based on deontic conditions. In deontic autonomy, an agent is permitted to do/decide/interpret/infer/etc. Not only is it practically able and in a condition to do so, but it can do this without violating a social or legal norm, or the user/designer prescriptions. As there are two kinds of autonomy there are two kinds of "empowerment" (giving autonomy): deontic empowerment versus practical, material empowerment [12]. Therefore, an additional dimension of *adjustment* should be taken into account, that is, the deontic one; in fact, not only practical autonomy but also deontic autonomy can be adjusted. The delegator (or the delegee) can attempt to modify (restrict or enlarge) either what the delegee is practically and actually able to do independent of the other, or what it is entitled to do. For example, when a delegee restarts negotiation—instead of directly modifying the task—it is implicitly asking some sort of permission or agreement. Obviously enough, in strong delegation (contract relation), the assignment of a task  $\tau$  to the delegee implicitly entails giving it the permission to do  $\tau$ . Adjusting the entitled space of freedom, or adjusting the practical space of freedom, is an interesting difference, but we cannot examine it in this paper. Notice that this theory would imply the same plan-based dimensions of delegation and help.

in  $\prod^d$  :  $\text{Dom-c}(\alpha_1 \alpha_2) = \text{True}$  if there is a set of rules  $(r_1, \dots, r_m)$  in  $\prod^d$  such that  $(\alpha_1 = \text{Lr}_1) \wedge (\alpha_2 \in \text{Rr}_m) \wedge (\text{Lr}_i \in \text{Rr}_{i-1})$  where  $\text{Lr}_j$  and  $\text{Rr}_j$  are, respectively, the left part and the right part of the rule  $r_j$  and  $2 \leq i \leq m$  (in the same way it is possible to define the  $\text{Dom-a}(\alpha_1 \alpha_2)$  operator considering the abstraction hierarchy rule set  $\prod^a$ ). We denote  $\prod_{\text{Ag}_x}$  as the  $\text{Ag}_x$ 's plan library, and  $\text{Act}_{\text{Ag}_x}$ , the set of actions known by  $\text{Ag}_x$ . The set of irreducible actions (through decomposition or specification rules) included in  $\prod_{\text{Ag}_x}$  is composed of two subsets: the set of actions that  $\text{Ag}_x$  believes to be elementary actions ( $\text{BAct}_{\text{Ag}_x}$ ) and the set of actions that  $\text{Ag}_x$  believes to be complex but for which it has no reduction rules ( $\text{NRAct}_{\text{Ag}_x}$ : *Non Reduced actions*). Then  $\text{BAct}_{\text{Ag}_x}$  is included in  $\text{Act}$  and possibly  $\text{BAct}_{\text{Ag}_x}$  is included or coincides with  $\text{BAct}$ . In fact, given an elementary action, an agent may (or not) know the body of that action. We define  $S_{\text{Ag}_x}$  as the *skill set* of  $\text{Ag}_x$ , the actions in  $\text{BAct}_{\text{Ag}_x}$  whose body is known by  $\text{Ag}_x$  (action repertoire of  $\text{Ag}_x$ ).<sup>5</sup> We call  $R$  the operator that, when applied to an action  $\alpha$ , returns the set of the *results* produced by  $\alpha$ .

2) *Definition of Delegation and Adoption*: The domain of MAS, collaboration [13], and teamwork are already familiar with the notion of delegation. However, our analysis is grounded on more basic notions [4]. In addition, our delegation theory is not limited to explaining and modeling interpersonal relationships; the basic concepts of our definition also apply to (and are necessary even if not sufficient for) other important concepts such as the following.

- *Institutional delegation*, in which the delegator transfers to the delegee not just some task but also some right, obligation, responsibility, power, and so on [12]. Of course, this notion is richer than our basic concept (see below).
- *Roles and prescriptions in organizations*; roles can be analyzed also as sets of delegated tasks [14].

In our model, *delegation and goal adoption are characterized in terms of the particular set of mental states (cognitive ingredients) of the agents involved in the interaction.*

Informally, *in delegation (reliance) an agent  $\text{Ag}_1$  needs or likes an action of another agent  $\text{Ag}_2$  and includes it in its own plan. In other words,  $\text{Ag}_1$  is trying to achieve some of its goals through  $\text{Ag}_2$ 's actions; thus  $\text{Ag}_1$  is relying upon the fact and has the goal that  $\text{Ag}_2$  performs a given action.  $\text{Ag}_1$  is constructing a multi-agent (MA) plan and  $\text{Ag}_2$  has a "part" in this plan:  $\text{Ag}_2$ 's task (either a state-goal or an action-goal).*

*In adoption (help) an agent  $\text{Ag}_2$  acquires and has a goal as (long as) it is the goal of another agent  $\text{Ag}_1$ , that is,  $\text{Ag}_2$  has the goal of performing an action because this action is included in the plan of  $\text{Ag}_1$ . So, also in this case,  $\text{Ag}_2$  plays a part in  $\text{Ag}_1$ 's plan (sometimes  $\text{Ag}_1$  has no plan at all but just a need, or a goal) since  $\text{Ag}_2$  is doing something for  $\text{Ag}_1$ .*

<sup>5</sup>In sum, an agent  $\text{Ag}_x$  has its own plan library  $\prod_{\text{Ag}_x}$  in which some actions ( $\text{CAct}_{\text{Ag}_x}$  and  $\text{NRAct}_{\text{Ag}_x}$ ) are complex actions (and it knows the reduction rules of  $\text{CAct}_{\text{Ag}_x}$ ) while some other actions ( $\text{BAct}_{\text{Ag}_x}$ ) are elementary actions (and it knows the body of a subset— $S_{\text{Ag}_x}$ —of them).

We consider the action/goal pair  $\tau = (\alpha, g)$  as the real object of delegation,<sup>6</sup> and we will call it a "task." Then by  $\tau$ , we will refer to the action ( $\alpha$ ), to its resulting world state ( $g$ ), or to both. We introduce an operator of delegation with three parameters:  $\text{Delegates}(\text{Ag}_1 \text{Ag}_2 \tau)$ , where  $\text{Ag}_1, \text{Ag}_2$  are agents and  $\tau = (\alpha, g)$  is the task. This means that  $\text{Ag}_1$  delegates the task  $\tau$  to  $\text{Ag}_2$ . In analogy with delegation we introduce the corresponding operator for adoption:  $\text{Adopts}(\text{Ag}_2 \text{Ag}_1 \tau)$ . This means that  $\text{Ag}_2$  adopts the task  $\tau$  for  $\text{Ag}_1$ :  $\text{Ag}_2$  helps  $\text{Ag}_1$  by caring about  $\tau$ .

3) *Dimensions of Delegation/Adoption*: We consider three main dimensions of delegation/adoption:

- 1) interaction-based;
- 2) specification-based;
- 3) control-based types of delegation/adoption [4].

Let us analyze these cases in more detail.

*Interaction-Based Types of Delegation*: Three general cases may be given: *weak, mild, and strong delegation*. They represent different degrees of strength of the established delegation. More precisely, we call *weak delegation* the delegation which is simply based on another's exploitation for the achievement of the task. In it there is no agreement, no request, or even (intended) influence:  $\text{Ag}_1$  is just exploiting in its plan a fully autonomous action of  $\text{Ag}_2$ . We call *mild delegation* that based on induction. In it there is no agreement, no request, but  $\text{Ag}_1$  is itself eliciting, and in some way inducing in  $\text{Ag}_2$  the desired behavior in order to exploit it. We call *strong delegation*, that based on explicit agreement, i.e., on the achievement by  $\text{Ag}_1$  of the task through an agreement with  $\text{Ag}_2$ .

More precisely, we introduce the *W-Delegates* operator for representing *weak delegation*. So the expression  $\text{W-Delegates}(\text{Ag}_1 \text{Ag}_2 \tau)$  represents the following *necessary mental ingredients*.

- a) The achievement of  $\tau$  is a *goal* of  $\text{Ag}_1$ .
- b)  $\text{Ag}_1$  believes that there exists another agent  $\text{Ag}_2$  that has the *power* of [8] achieving  $\tau$ .
- c)  $\text{Ag}_1$  believes that  $\text{Ag}_2$  will achieve  $\tau$  in time and by itself (without  $\text{Ag}_1$ 's intervention). (If  $\text{Ag}_2$  is a cognitive agent,  $\text{Ag}_1$  believes that  $\text{Ag}_2$  *intends* to achieve  $\tau$ ).
- d)  $\text{Ag}_1$  *prefers*<sup>7</sup> to achieve  $\tau$  through  $\text{Ag}_2$ .
- e) The achievement of  $\tau$  through  $\text{Ag}_2$  is the choice (goal) of  $\text{Ag}_1$ .
- f)  $\text{Ag}_1$  has the goal [*relativized* [15] to e)] of not achieving  $\tau$  by itself.

We consider a), b), c), and d) what the agent  $\text{Ag}_1$  views as a "*potential for relying on*" agent  $\text{Ag}_2$ , its *trust* in  $\text{Ag}_2$ ; and [e) and f)] what  $\text{Ag}_1$  views as the "*decision to rely on*"  $\text{Ag}_2$ . We consider "potential for relying on" and "decision to rely on" as

<sup>6</sup>We assume that *delegating an action necessarily implies delegating some result of that action* (i.e., expecting some results from  $\text{Ag}_2$ 's action and relying on it for obtaining those results). Conversely, *to delegate a goal state always implies the delegation of at least one action (possibly unknown to  $\text{Ag}_1$ ) that produces such a goal state as a result* (even when  $\text{Ag}_1$  asks  $\text{Ag}_2$  to solve a problem, to bring it about that  $g$  without knowing or specifying the action,  $\text{Ag}_1$  necessarily presupposes that  $\text{Ag}_2$  should and will do some action and relies on this).

<sup>7</sup>This means that  $\text{Ag}_1$  believes that either the achievement of  $\tau$  or a broader goal  $g'$  that includes the achievement of  $\tau$ , implies  $\text{Ag}_1$  to be dependent on  $\text{Ag}_2$  [6], [16]. Moreover, d) implies  $\text{Ag}_1$ 's goal that  $\text{Ag}_2$  achieves  $\tau$ .

two constructs temporally and logically related to each other (for a weak delegation example, see Ex1 in the Appendix).<sup>8</sup>

We introduce the *M-Delegates* operator for representing *mild delegation* (Ex5 in Appendix).  $M\text{-Delegates}(Ag_1 Ag_2 \tau)$  represents the following necessary mental ingredients:

$a' \equiv a; b' \equiv b; d' \equiv d; e' \equiv e; f' \equiv f;$

$c')$   $Ag_1$  does not believe that  $Ag_2$  will achieve  $\tau$  by itself (without  $Ag_1$ 's intervention).

$g')$   $Ag_1$  believes that if  $Ag_1$  realizes an action  $\alpha'$  then it is more probable that  $Ag_2$  intends  $\tau$ . But  $Ag_2$  does not adopt  $Ag_1$ 's goal that  $Ag_2$  intends  $\tau$ .

$h')$   $Ag_1$  intends to do  $\alpha'$  relativized to  $(e')$ .

We consider  $a')$ ,  $b')$ ,  $c')$ ,  $d')$ , and  $e')$  what agent  $Ag_1$  views as a “*potential for relying on*” agent  $Ag_2$ ; and  $f')$ ,  $g')$ , and  $h')$  what  $Ag_1$  views as the “*decision to rely on*”  $Ag_2$ .<sup>9</sup>

We introduce the *S-Delegates* operator for representing *strong delegation* (Ex6 in Appendix). So the expression  $S\text{-Delegates}(Ag_1 Ag_2 \tau)$  represents the following *necessary* mental ingredients:

$a'' \equiv a; b'' \equiv b; c'' \equiv c; d'' \equiv d; e'' \equiv e; f'' \equiv f;$

$g'')$   $Ag_1$  believes that if  $Ag_1$  realizes an action  $\alpha'$  there will be this result:  $Ag_2$  will intend  $\tau$  as the consequence of the fact that  $Ag_2$  adopts  $Ag_1$ 's goal that  $Ag_2$  intends  $\tau$  (in other words,  $Ag_2$  will be socially committed to  $Ag_1$ ).

$h'')$   $Ag_1$  has the goal (*relativized* to  $(e'')$ ) of not achieving  $\tau$  by itself.

We consider  $a'')$ ,  $b'')$ ,  $c'')$ ,  $d'')$ , and  $e'')$  what agent  $Ag_1$  views as a “*potential for relying on*” agent  $Ag_2$ ; and  $f'')$ ,  $g'')$ , and  $h'')$  what  $Ag_1$  views as the “*decision to rely on*”  $Ag_2$ .

For a corresponding analysis of adoption, and for how the kind of interaction between client and contractor influences the adoption itself, see [4].

*Specification-based types of delegation/adoption.* How is the task specified in delegation and how does this specification influence the contractor's autonomy? The object of delegation/adoption ( $\tau$ ) can be minimally specified (*open delegation*), completely specified (*close delegation*), or specified at any intermediate level. Let us consider two cases:

- i) *Pure executive (close) delegation:* Here, the client (or the contractor) believes it is delegating (adopting) a completely specified task; what  $Ag_1$  expects from  $Ag_2$  is just the execution of a sequence of elementary actions (or what  $Ag_2$  believes  $Ag_1$  delegated to it is simply the ex-

<sup>8</sup>With respect to weak delegation, it is interesting to analyze the possibilities of the  $Ag_2$ 's mind. We should distinguish between two main cases:  $Ag_2$  knows  $W\text{-Delegates}(Ag_1 Ag_2 \tau)$  and  $Ag_2$  does not know  $W\text{-Delegates}(Ag_1 Ag_2 \tau)$ . In other words, a weak delegation is possible even if the delegee knows it. Either this knowledge has no effect (the achievement of  $Ag_1$ 's goal is just a side-effect known by  $Ag_2$ ) (see Ex2 in Appendix) or this knowledge changes  $Ag_2$ 's goal:  $Ag_2$  can either arrive to a spontaneous and unilateral help or to a reactive, hostile attitude (Ex3 and Ex4 in Appendix).

<sup>9</sup>In analogy with what we have said in weak delegation, in mild delegation, we should distinguish between two main cases about the possible mental states of  $Ag:Ag$  knows  $M\text{-Delegates}(Ag_1 Ag_2 \tau)$ , and  $Ag_2$  does not know  $M\text{-Delegates}(Ag_1 Ag_2 \tau)$ . So, it is possible to have a mild delegation even if the delegee knows it and if, consequently, it changes its own behavior to favor or to hamper its success.

ecution of a sequence of elementary actions) [Ex9 in Appendix].<sup>10</sup>

- ii) *Open delegation:* When the client (contractor) believes it is delegating (adopting) a noncompletely specified task: either  $Ag_1$  ( $Ag_2$ ) is delegating (adopting) an abstract action, or it is delegating (adopting) just a result (i.e., a state of the world).<sup>11</sup>  $Ag_2$  can realize the delegated (adopted) task by exerting its autonomy. We can have several possible level of openness of the delegated (adopted) task (Ex10 in Appendix).

*Control-based types of delegation.* We view control as a (meta) action aimed at

- a) ascertaining whether another action has been successfully executed or if a given state of the world has been realized or maintained (*feedback*, checking);
- b) dealing with the possible deviations and unforeseen events in order to positively cope with them (*intervention*).

When the client is delegating a given *object-action*, what about its control actions? Considering, for example, that the control action is executed by a single agent, if  $\text{Delegates}(Ag_1 Ag_2 \tau)$  there are at least four possibilities:

- i)  $Ag_1$  delegates the control to  $Ag_2$ : the client does not (directly) verify the success of the delegation to the contractor;
- ii)  $Ag_1$  delegates the control to a third agent;
- iii)  $Ag_1$  gives up the control: nobody is delegated to control the success of  $\alpha$ ;
- iv)  $Ag_1$  maintains the control for itself.

Each of these possibilities could be explicit or implicit in the delegation of the action, in the roles of the agents (if they are part of a social structure), in the previous interactions between the client and contractor, etc.

In Fig. 1, we summarize the three main dimensions of delegation: each characterizes the variability of delegation action. The delegee's autonomy increases toward the origin of the Cartesian space within the solid. Each of these dimensions implies, in fact, a specific aspect of the delegee's autonomy about the task.

### III. ADJUSTMENT OF DELEGATION/ADOPTION

#### A. Some Pertinent Distinctions

1) *Run-Time Adjustment:* We can consider the adjustment of autonomy (the revision of delegation) in three different time periods (Fig. 2):

- i) *after the delegation event, but before the execution of the task;*
- ii) *run-time, with work-in-progress;*
- iii) *at the end of the performance* and the realization of the task; in this case the adjustment of autonomy will have an effect only on a future delegation (*learning*).

<sup>10</sup>More formally, either  $\alpha \in S_{Ag_1}$ ,  $\alpha \in BAct_{Ag_1}$  ( $\alpha \in S_{Ag_2}$ ,  $\alpha \in BAct_{Ag_2}$ ), or  $g$  is the relevant result of  $\alpha$ , and  $\alpha \in S_{Ag_1}$  or  $\alpha \in BAct_{Ag_1}$  ( $\alpha \in S_{Ag_2}$ ,  $\alpha \in BAct_{Ag_2}$ ).

<sup>11</sup>More formally, either  $\alpha \in CAct_{Ag_1}$ , or  $\alpha \in NAct_{Ag_1}$  (either  $\alpha \in CAct_{Ag_2}$ , or  $\alpha \in NAct$ ), and when  $g$  is the relevant result of  $\alpha$  and  $\alpha \in CAct_{Ag_1}$  or  $\alpha \in NAct_{Ag_1}$  ( $\alpha \in CAct_{Ag_2}$ , or  $\alpha \in NAct_{Ag_2}$ ).

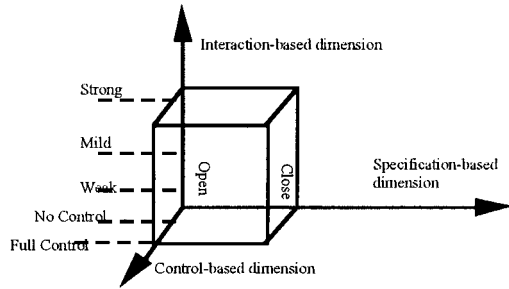


Fig. 1. Three main dimensions of delegation.

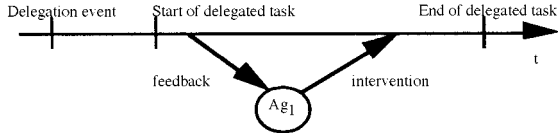


Fig. 2. Feedback and intervention after delegation.

We will focus here on *run-time* adjustment, which is what really matters in human-machine interaction and in MA cooperation, and call it simply *adjustment*.

We will first examine the problem of adjustable autonomy *in a broad sense*, i.e., as adjusting the level and the kind of delegation/adoption (in our terminology *delegation conflicts* [17]). We claim that this is the right theoretical frame for understanding also the adjustment of autonomy in a strict sense, since *any autonomy adjustment requires a delegation adjustment* but not vice versa (see Section III-B).

In Sections V-A-D we will analyze the general reasons for delegation/adoption adjustment. Let us here consider the taxonomy of the adjustments (some of which will be neglected because they are meaningless), their nature, and their importance. Each of the possible adjustments is *bilateral*, i.e., either the client or the contractor can try to modify the previous delegation.

### B. Delegation/Adoption Adjustments With Respect to the Interaction Dimension

As described in Table I there are (with respect to the interaction dimension) several possibilities of adjustment; they are determined by

- 1) the *agent* who has the initiative of the adjustment;
- 2) the *starting state* (the kind of delegation or adoption acting in that given instant and that the *agent* intends to modify);
- 3) the *final state* (the kind of delegation or adoption to which the *agent* intends to arrive).

A few cases deserve some comments.

Line 1: This can be inferred from the difference between the mental ingredients of weak and mild delegation: in fact,  $c$  is replaced by  $c'$ ,  $g'$ , and  $h'$ . In other words,  $Ag_1$  does not believe that  $Ag_2$  will achieve  $\tau$  without any influence and so decides to realize an action  $\alpha'$  that could produce this influence. In this case there is still no social commitment [18] by  $Ag_2$ :  $Ag_2$  does not adopt  $Ag_1$ 's goal that  $Ag_2$  intends  $\tau$  (Ex11 in the Appendix).

TABLE I  
ADJUSTMENTS WITH RESPECT TO THE INTERACTION DIMENSION

Line Number	Agent that has the initiative of the adjustment	Starting State	Final State
1	Delegator	Weak delegation	Mild delegation
2	Delegator	Weak delegation	Strong delegation
3	Delegator	Mild delegation	Strong delegation
4	Delegator	Mild delegation	Weak delegation
5	Delegator	Strong delegation	Weak delegation
6	Delegator	Strong delegation	Mild delegation
7	Deleegee	Weak delegation	Mild delegation
8	Deleegee	Weak delegation	Strong delegation
9	Deleegee	Mild delegation	Strong delegation
10	Deleegee	Strong delegation	Mild delegation
11	Deleegee	Strong delegation	Weak delegation
12	Deleegee	Mild delegation	Weak delegation
13	Delegator	Weak adoption	Strong adoption
14	Delegator	Strong adoption	Weak adoption
15	Deleegee	Weak adoption	Strong adoption
16	Deleegee	Strong adoption	Weak adoption

Line 2: This is where  $g''$  and  $h''$  are added beliefs. In other words,  $Ag_1$  tries to achieve  $\tau$  through a social commitment of  $Ag_2$ : for this it realizes  $\alpha'$  (Ex12 in the Appendix).

Lines 8 and 9: These could represent the willingness of  $Ag_2$  to convert  $Ag_1$ 's exploitation into a clear social commitment between them (Ex13 in Appendix).

Lines 13 and 14: These are linked with the initiative of  $Ag_1$  in the case in which  $Ag_1$  is aware of  $Ag_2$ 's adoption (Ex14 in the Appendix).

### C. Delegation/Adoption Adjustments with Respect to the Specification Dimension

Also in these cases, we must consider the intervention of both the client and the contractor with regard to delegation and adoption, respectively.

1) *Deleegee's Adjustments*: Suppose that  $Delegates(Ag_1, Ag_2, \tau)$  and  $\tau$  are included in a more general  $Ag_1$ 's plan aimed at achieving goal  $g'$  through a complex action  $\alpha'$ . Moreover,  $Dom-c(\alpha', \alpha), \tau = (\alpha, g)$ , and  $\tau' = (\alpha', g')$ . We have three main delegees' adjustments.

a) *Reduction of Help*: Here the deleegee provides less help on  $\tau$  than delegated. If  $Adopts(Ag_2, Ag_1, \tau_1)$  AND  $Dom-c(\alpha, \alpha_1)$  with  $\tau_1 = (\alpha_1, g_1)$ , the deleegee reduces the task to a subpart of the requested one (Ex16 in the Appendix).

b) *Extension of Help*: Here the deleegee provides more help on  $\tau$  than delegated. If  $Adopts(Ag_2, Ag_1, \tau_1)$  AND  $Dom-c(\alpha_1, \alpha)$  AND  $(Dom-c(\alpha', \alpha_1))$  OR  $\alpha' \equiv \alpha_1$  with  $\tau_1 = (\alpha_1, g_1)$ ; the deleegee goes beyond what has been delegated by the client without changing the delegator's plan. In fact, the deleegee chooses a task that satisfies a higher level task (within the general delegator's intended plan) compared with the delegated task (Ex17 in the Appendix).

c) *Critical Help*: It is the case in which the deleegee provides a qualitatively different action/help than what is delegated. Let us look at some subcases.

- *Simple Critical Help*: If  $Adopts(Ag_2, Ag_1, \tau_x)$  AND  $g \in R(\alpha_x)$  with  $\tau_x = (\alpha_x, g)$ ; the deleegee achieves the goal(s) of the delegated plan/action, but it changes that plan/action.
- *Critical Overhelp*: If  $Adopts(Ag_2, Ag_1, \tau_x)$  AND  $g_1 \in R(\alpha_x)$  AND  $Dom-c(\alpha_1, \alpha)$  AND  $(Dom-c(\alpha', \alpha_1))$  OR  $\alpha' \equiv \alpha_1$  with  $\tau_1 = (\alpha_1, g_1)$ ,  $\tau_x = (\alpha_x, g_1), R(\alpha_x)$  the set of the results produced

by  $\alpha_x$ ; the delegee implements both a simple critical help and an extension of help (it chooses a task that satisfies a higher level task with respect to the task delegated and achieves the goal(s) of this higher task, while changing the expected plan) (Ex18 in the Appendix).

- *Hypercritical Help* If  $\text{Adopts}(\text{Ag}_2 \text{ Ag}_1 \tau_1)$  AND  $(g_1 \neq g')$  AND  $(g_1 \neq g)$  AND  $(g_1 \in I_{\text{Ag}_1})$  with  $\tau_1 = (\alpha_1, g_1)$ , and  $I_{\text{Ag}_1}$  is the set of the interests of  $\text{Ag}_1$ .  $\text{Ag}_2$  adopts goals or interests of  $\text{Ag}_1$  that  $\text{Ag}_1$  itself did not take into account; by doing so,  $\text{Ag}_2$  neither performs the delegated action/plan nor achieves the results that were delegated.

2) *Delegator's Adjustment*: Suppose that  $\text{Delegates}(\text{Ag}_1 \text{ Ag}_2 \tau)$ , and that  $\text{Ag}_1$  intends to change that delegation. Suppose also that  $\text{Ag}_1$  is achieving goal  $g'$  through plan  $\tau'$ , with  $\text{Dom-c}(\alpha' \alpha)$ . We can have five main delegator's adjustments.

a) *Reduction of Delegation*: It is the case in which there is a new delegation:  $\text{Delegates}(\text{Ag}_1 \text{ Ag}_2 \tau_1)$  AND  $\text{Dom-c}(\alpha' \alpha_1)$  with  $\tau_1 = (\alpha_1, g_1)$ , the delegator adjusts the original delegation, by reducing the task that the contractor must realize (the client reduces the task to a subpart of the previous requested task).

b) *Extension of Delegation*:  $\text{Delegates}(\text{Ag}_1 \text{ Ag}_2 \tau_1)$  AND  $\text{Dom-c}(\alpha_1 \alpha)$  AND  $(\text{Dom-c}(\alpha' \alpha_1) \text{ OR } \alpha' \equiv \alpha_1)$  with  $\tau_1 = (\alpha_1, g_1)$ , the delegator adjusts its delegation in such a way that its new request goes beyond what has been originally delegated without changing the previous plan.

c) *Modification of Delegation*: In analogy with the delegee's adjustments, it consists of four subcases (modification of the previous delegated task just changing the previous goal; modification of the previous delegated task considering an overgoal and changing the plan to obtain that overgoal; modification of the previous delegated task considering a subgoal and changing the plan to obtain that subgoal; modification of the previous delegated task changing both the plan and the goal). In this paper, for reasons of brevity, we do not analyze them.

d) *Openness of Delegation*: If  $\text{Delegates}(\text{Ag}_1 \text{ Ag}_2 \tau_x)$  AND  $\text{Dom-a}(\alpha_x \alpha)$  in words, the delegator adjusts its own delegation so that the new delegated plan is more abstract.

e) *Closing of Delegation*: If  $\text{Delegates}(\text{Ag}_1 \text{ Ag}_2 \tau_x)$  AND  $\text{Dom-a}(\alpha \alpha_x)$  in words, the delegator adjusts its own delegation so that the new delegated plan is more specified.

In these cases in which the client has to adjust the delegee's adoption, we have the same typology of the case a), b), c), d) and e) in which the delegator was changing a previous delegation); the difference now is that the adjustment must be "requested" (through an order, a request, negotiation, influence, induction, etc.) to the contractor.

#### D. Channels for the Bilateral Adjustment

Channels and protocols are necessary for adjusting delegation and help. On the user/client's side, they are useful for monitoring (reporting, observing, inspecting) and intervention (instructions, guidance, helps, repair, brake); on the delegee's side it is useful to have some room for discretion and practical innovation. For both client and contractor, are useful channels and protocols for communication and renegotiation during role-playing and task execution.

	Feedback from the	
	Client	Contractor
Feedback to the	Client	Inspection Surveillance
	Contractor	Guidance
		Report
		Self Monitoring

Fig. 3. How feedback determines different kinds of control.

*The Client's Side*:  $\text{Ag}_1$  can adjust run-time on its own delegation to  $\text{Ag}_2$  (and possibly  $\text{Ag}_2$ 's autonomy) if it is in the condition of

- a) receiving on time the necessary information about  $\text{Ag}_2$ 's performance (*feedback*);
- b) intervening on  $\text{Ag}_2$ 's performance to change it before its completion (*intervention*).

In other words,  $\text{Ag}_1$  must have some form of *control* on and during  $\text{Ag}_2$ 's task realization. *Control* requires feedback plus intervention.<sup>12</sup> Otherwise no adjustment is possible. Obviously, the feedback, i.e., monitoring, useful for a run-time adjustment must be provided timely for the intervention. In general, the feedback activity is the precondition for an intervention; however, it is also possible that either only the feedback or only the intervention hold. In sum, in order to guarantee agents a dynamic adjustment of delegation and their mixed initiative in MAS we have to provide such an *infrastructure*, while in human computer interaction we have to provide the user with those two channels.

In the case in which  $\text{Ag}_2$  has the initiative (it is starting with an adoption action), if  $\text{Ag}_1$  wants to change this adoption it needs a *communication channel* with  $\text{Ag}_2$ .

*The Delegee's Side*:  $\text{Ag}_2$  can run-time adjust the delegation of  $\text{Ag}_1$  (and possibly its own autonomy) if it is in the condition of either:

- a) having a *communication channel* for (*restarting negotiation* by offering/proposing a different level of help to  $\text{Ag}_1$ ;
- b) having enough practical freedom to *directly change* the action (this condition should be by definition a characteristic of autonomous agents; otherwise, they should not be considered agents but simply tools).

Delegees should not necessarily negotiate or give advice in order to change their delegated tasks; they might have a full initiative. This entails a metalevel autonomy. Of course, the delegee must also have feedback about its own execution, but this is true in general for goal-directed actions and for agenthood.

To sum up, if an agent has the initiative of a delegation/adoption then, in order to adjust that initiative, it is not obliged to communicate with the other agent.

As for the necessary *feedback for adjustment* we can distinguish among: inspection, report, guidance, and self-monitoring (Fig. 3).

We consider five kinds of intervention action:

- *stopping the task* (the delegation or the adoption process is suddenly interrupted);
- *substitution* (an intervention allocates (part of) the task to the intervening agent);

<sup>12</sup>We will call *control activity* the functional combination of two more specific activities: monitoring and intervention.

TABLE II  
DIFFERENT KINDS OF CLIENT INTERVENTION

	Client's message	Client's direct action
Stopping the task	Stop	Stopping intervention
Substitution	I do it	It realizes an action of the task
Correction of delegation	Change that action with this other	It introduces constraints such that an action is changed with another
Specification of delegation	Make that plan in this way	It introduces constraints such that a plan is specified
Repairing of delegation	Add this action to the task	It introduces constraints such that a new action must be realized to favor success of the task

- *correction of delegation* (after the intervention, the task is partially or totally changed);
- *specification of delegation* (after the intervention, the task is more constrained);
- *repairing of delegation* (the intervention leaves the task activity unchanged but it introduces new actions necessary to achieve the goal(s) of the task itself).

Each of these interventions could be realized through either a *communication act* or a *direct action* on the task by the intervening agent (Table II).

#### E. Protocols for Control

Starting from our model, it is possible to identify some guidelines for designing interaction in human machine interaction (HMI) or in MAS. For example, our model makes it clear that agent's control requires either communication or simple action–perception loop. On the one side,  $Ag_1$  can monitor  $Ag_2$  without any communication (without sending any specialized message or signal) by simply observing it; and  $Ag_1$  can also influence  $Ag_2$  by physical/practical actions and interventions on  $Ag_2$  or on the world. For example,  $Ag_1$  can brake and stop  $Ag_2$ . On the other side,  $Ag_1$  can monitor  $Ag_2$  thanks to messages sent by  $Ag_2$  to  $Ag_1$  (reports), and can influence  $Ag_2$  by sending it messages (instructions, warnings, etc.)

In the classical contract net a few kinds of messages (and only messages) aimed at this. There was the possibility of “reports” from the delegated agents, and messages to inform about the abort of the collaboration or about its successful completion. A much richer variety of interaction protocols for control (messages and actions) is needed. Let us provide some examples from the controller's point of view.

##### 1) Examples for the Monitoring:

PRAXIS:

- a) *inspection*: (visiting  $Ag_2$  for ascertaining if everything is as expected);
- b) *internal inspection*: (inspecting some inside data of the agent for checking its reasoning, agenda, plan library, plan, etc.);
- c) *surveillance*: (by sample) observing  $Ag_2$ 's activity and partial results and the environment for avoiding damages;
- d) *detecting*: analyzing some traces of  $Ag_2$ 's activity in order to (abductively) check whether its behavior has been correct and at what stage it is.

COMMUNICATION:

- a) *report requests* (“let me know what is happening”; “any news?”);
- b) *inspective questions* (“is everything as expected?” “what are you doing?” “is  $p$  true?”).

##### 2) Examples for the Intervention:

PRAXIS:

- 1) *substitution*:  $Ag_1$  performs (part of) an action previously allocated to  $Ag_2$ ;
- 2) *support*:  $Ag_1$  modifies the conditions of the world so that  $Ag_2$  can successfully perform its action or damages can be prevented;
- 3) *brake*:  $Ag_1$  stops  $Ag_2$ 's activity (either by external obstacles or directly acting upon/in  $Ag_2$ 's body or software);
- 4) *tuning*:  $Ag_1$  modifies and correct  $Ag_2$ 's action (either by external obstacles or directly acting upon/in  $Ag_2$ 's body or software);
- 5) *repair*:  $Ag_1$  acts in order to repair damages of  $Ag_2$ 's action and to recover from failures.

COMMUNICATION:

- 1) *alert/warning*:  $Ag_1$  alerts  $Ag_2$  about unexpected events or possible danger;
- 2) *advice*:  $Ag_1$  provides  $Ag_2$  with some possible recipe, solution, or better action just as an advice ( $Ag_2$  is free to accept or not);
- 3) *instructions*:  $Ag_1$  gives instruction to  $Ag_2$  about how to proceed ( $Ag_1$  is specifying the previously “open” delegation);
- 4) *threats*:  $Ag_1$  threatens  $Ag_2$  to induce it to do what  $Ag_2$  should do;
- 5) *reminding*:  $Ag_1$  reminds  $Ag_2$  about what it should do;
- 6) *stop*:  $Ag_1$  orders  $Ag_2$  to stop its activity;
- 7) *abort*:  $Ag_1$  estinguishes its delagation to  $Ag_2$ .

We are working on a model of when and why a given form of intervention (e.g., to stop  $Ag_2$ ) is useful or better than others; and what kind of feedback (e.g., surveillance versus report) is appropriate for a given task level of trust and possible kind of intervention.

#### IV. FROM DELEGATION ADJUSTMENT TO AUTONOMY ADJUSTMENT

As we said, a delegation adjustment does not always produce a change in the delegee's autonomy (by limiting, restricting, or vice versa, enlarging or expanding it). The main causes of autonomy adjustment are the following.

- there is a change of  $Ag_2$ 's entitlement at the metalevel ( $Ag_2$  can refuse, negotiate, change the delegation); or it is not entitled but in fact takes such an initiative (*meta-autonomy adjustment*);
- the new task is more or less *open* than the former (*realization-autonomy adjustment*);
- there is more or less control on  $Ag_2$  (*control-dependent autonomy adjustment*);

TABLE III  
HOW AUTONOMY CHANGES WHILE ADJUSTING DELEGATION AND HELP

	Meta Autonomy	Realization Autonomy
Reduction of help	Increased	Equal
Extension of help	Increased	Increased or Equal
Critical help	Increased	Increased or Reduced or Equal
Reduction of delegation	Equal	Equal or Reduced
Modification of delegation	Equal	Increased or Equal
Critical delegation	Equal	Increased or Reduced or Equal
Openness of delegation	Equal	Increased
Closing of delegation	Equal	Reduced

- there is a change in the strength of the delegation (*interaction-dependent autonomy adjustment*).

Each of these autonomy adjustments can be *bilateral* (realized by either the client or the contractor or both) and *bidirectional* (either augmenting or reducing the autonomy itself).

#### A. Adjusting Meta-Autonomy and Realization Autonomy of the Deleegee

By crossing the two first kinds of adjustment (meta-autonomy adjustment and realization-autonomy adjustment) with the delegation adjustments (D), we obtain the results shown in Table III: rows 1–3 show the adjustments of delegation by the deleegee (deleegee’s adjustments) while rows 4–8 show the adjustments by the client (client’s adjustments) on its own previous delegation. In particular, we can see that:

- *when there is a deleegee’s adjustment there is always a change of its meta-autonomy* (the deleegee decides to change the client’s delegation) while there is not always a change in its realization autonomy. For example, in the *reduction of help*, realization autonomy remains the same because the deleegee realizes just a part of the delegated task (but this part was also included in the previously delegated task). In other words, the deleegee does not have more autonomy as for how to realize  $\tau$ . Conversely, in the *extension of help*, there are two possibilities: 1) the deleegee has more realization autonomy when the adopted plan includes some (not delegated) part which is not completely specified (thus, the deleegee has more discretion in its realization) and 2) the deleegee has the same realization autonomy if the adopted plan does not need more discretion than the delegated one. Finally, in *critical help*, given its possibility to choose new actions, there is always more realization autonomy.
- *When there is a client’s adjustment the deleegee’s meta-autonomy never changes* (in fact, the client itself takes the initiative to modify the delegation). As for the deleegee’s realization autonomy we can say that: in the *reduction of delegation* case,  $Ag_2$ ’s autonomy of execution (if its discretionary power is reduced with the new delegation) is reduced or it remains unchanged (suppose that the old task was completely specified in all details). In the *extension of delegation* case, either the autonomy of realization increases (if

the new task presupposes some action—not included in the old one—with a certain degree of openness) or it remains unchanged (if this new task was completely specified in all details). In the *critical delegation* case, the autonomy of realization of the deleegee increases or not depending on whether respectively the new actions are more or less open than the old ones. In the *openness of delegation* case, the autonomy of realization of the deleegee always increases (openness is in fact a factor that increases the discretion of the deleegee). Vice versa, in the case of *closing of delegation*, the deleegee’s autonomy of realization is always reduced.

#### B. Adjusting Autonomy by Modifying Control

As already observed a very important dimension of autonomy is the control activity of the adopted/delegated task. Given that control is composed by feedback plus intervention, adjusting it means to adjust (at least one of) its components.

1) *Adjusting the Frequency of the Feedback*: The *frequency of the feedback on the task* can be

- *purely temporal* (when the monitoring (*mon*) or the reporting (*rep*) is independent of the structure of the activities in the task, they only depend on a temporal choice);
- *linked with the task phases* (when the activities of the task are divided in phases and the *mon* or the *rep* is connected with them).

Client and contractor can adjust the frequency of their feedback activity in three main ways:

- by *changing the temporal intervals* fixed at the beginning of the task delegation or task adoption (when the *mon/rep* is purely temporal);
- by *changing the task phases* in which the *mon/rep* is realized with respect to those fixed at the beginning of the task delegation;
- by *moving from* the purely temporal *mon/rep* to the task phases *mon/rep* (or vice versa).

2) *Adjusting the Frequency and Kind of Intervention*: As explained above, the intervention is strictly connected with the presence of the *mon/rep* on the task, even if, in principle, both the intervention and the *mon/rep* could be independently realized. In addition, also the occurrence of intervention and *mon/rep* are correlated. More precisely, the intervention can occur:

- never*;
- just sometimes* (during some phase or at specified times; a special case of this is at the end of the task);
- at any phase or at any time (depending on the necessity)*.

The adjustment of the frequency of intervention by the client is an important case of adjustment of the contractor’s autonomy. Suppose that at the beginning there is an agreement about the fact that the established frequency of intervention is *never*, and suppose that the client intervenes one or more times during the contractor’s task realization: the contractor’s autonomy has been reduced. In general, a contractor is more autonomous if the frequency of the client’s intervention is low. So the adjustments by the client in this direction (low frequency of interventions)

produce an increase in contractor's autonomy. If the client adjusts the possible kind of intervention established at the beginning of delegation it could increase or reduce the contractor's autonomy depending on this adjustment.

## V. WHEN TO ADJUST THE AUTONOMY OF THE AGENTS

We will examine in this section the general principles (criteria) for adjusting (restricting or expanding) the delegee's autonomy by both the user/client and the delegee.

### A. When to Reduce the Delegee's Autonomy

In this preliminary identification of reasons for autonomy adjustment, we prefer a more qualitative and simplified view, not necessarily related with a probabilistic framework like the one we will use in Section VI-A. Of course, to be more precise, one should specify that what changes is the subjective probability assigned to those events (beliefs). For example, at the time of the delegation,  $Ag_1$  believed that the probability of  $Ag_2$ 's mistakes was  $p$  (and this expectation was compatible with the decision of delegating a given degree of autonomy), while now  $Ag_1$  realizes that this probability has changed (higher or lower than expected).

Let us simplify the issue in the following schema.

Reduce the delegee's autonomy WHEN (classes of reasons)

- $Ag_1$  believes that  $Ag_2$  is not doing (in time) what  $Ag_1$  has delegated to it;
- $Ag_1$  believes that  $Ag_2$  is working badly and makes mistakes (because of lack of competence, knowledge, control, etc.);
- $Ag_1$  believes that there are unforeseen events, external dangers and obstacles that perhaps  $Ag_2$  is not able to deal with;
- $Ag_1$  believes that  $Ag_2$  is going beyond its role or task, and  $Ag_1$  is not happy about this (because of lack of trust or of conflict of power).<sup>13</sup>

THEN (reduction of autonomy):

$Ag_1$  will reconsider its delegation to  $Ag_2$ , and  $Ag_2$ 's level of autonomy in order to reduce it by either specifying the plan (task) or by introducing additional control, or constraining the interaction (strong delegation), etc.

### B. When to Expand the Delegee's Autonomy

WHEN (classes of reasons):

- $Ag_1$  believes that  $Ag_2$  is doing or can do better than previously expected (predicted);
- $Ag_1$  believes that the external conditions are more favorable than expected;
- $Ag_1$  believes that  $Ag_2$  is working badly and makes mistakes (because of lack of flexibility or because of too much control, etc.);
- $Ag_1$  believes that  $Ag_2$  can do more than previously assigned, or can find its own situated way of solving the problem

THEN (expansion of autonomy):

$Ag_1$  will change the delegation to  $Ag_2$  and  $Ag_2$ 's level of autonomy in order to expand it by either letting the plan (task) less specified or reducing control or making the interaction weaker, etc.

### C. When to Limit One's Own Autonomy

Let us now consider some (collaborative) reasons of adjustment on the delegated agent's side.

WHEN (classes of reasons):

- $Ag_2$  comes to believe that it is not able to do all the task (level of self-confidence);
- $Ag_2$  comes to believe that there are unforeseen events, external dangers, and obstacles that it is not able to deal with.

THEN (reduction of autonomy):

$Ag_2$  will reconsider the received delegation (for example, providing subhelp and doing less than delegated) and its level of autonomy in order to reduce it by either asking for some specification of the plan (task) or for the introduction of additional control, e.g., "give me instructions, orders; monitor, help, or substitute me."

### D. When to Expand One's Own Autonomy

WHEN (classes of reasons):

- $Ag_2$  gets to a grounded belief that it is able or in condition of doing more or providing a better solution for the delegated goal (within  $Ag_1$ 's plan or also with regard to  $Ag_1$ 's other desires and interests);
- it is not forbidden or it is (explicitly or implicitly) permitted by  $Ag_1$  that  $Ag_2$  takes such a collaborative initiative;
- $Ag_2$  believes that  $Ag_1$  will accept and enjoy its initiative (because convenience largely exceeds surprise or distress).

THEN (expansion of autonomy):

$Ag_2$  will reconsider the received delegation and level of autonomy in order to go beyond those limits by directly providing, for example, over-help or critical-help (doing more and better).

(When the second and third conditions above are not realized,  $Ag_2$  could take the initiative of communicating for offering the new solution or asking for a permission, and in fact for renegotiating the delegation.)

## VI. TRUST AS THE COGNITIVE BASIS FOR ADJUSTING AUTONOMY

What we have just seen (principles and reasons for bilateral delegation and autonomy adjustment) can be considered from another perspective; in fact, it is just a matter of trust.

Trust, being the mental ground and counterpart of delegation, plays a major role in adjustment: limiting autonomy is usually due to a trust crisis while expanding autonomy is usually due to an increased trust.

### A. Client's Perspective: Revising Trust and Adjusting Delegation and Autonomy

In our view [5], trust is a structured set of intensional attitudes characterizing the mind of a truster. These attitudes mainly

<sup>13</sup>Notice that in all those cases the delegator's expectations on which trust and reliance were based are disappointed.

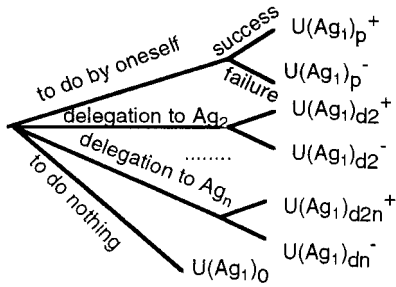


Fig. 4. Scenario of the possible choices.

consist of beliefs and goals; in particular, suppose there are two agents  $Ag_1$  and  $Ag_2$  and a task  $\tau$  and  $B_{Ag_1,t}$  and  $G_{Ag_1,t}$  are, respectively,  $Ag_1$ 's set of beliefs and goals at time  $t$ . We will call *trust-attitudes* the operator that, when applied to two agents ( $Ag_1$  and  $Ag_2$ ), a task ( $\tau$ ), and a temporal instant ( $t$ ), returns the set of beliefs and goals (of  $Ag_1$  on  $Ag_2$  about  $\tau$  at the time  $t$ ) useful to a trust relation. In particular, we propose [5] a set of mental ingredients—on which the trust relation is based—constituted by two subsets:

- $\Omega$ ) a set of mental states ( $MS-C_{T_{Ag_1,Ag_2,t}}$ )—called *core trust*—with these components:
  - $\Omega i$ ) a set of  $Ag_1$ 's *goals* ( $G_{Ag_1,t}$ ) in order to trust  $Ag_2$ ;
  - $\Omega ii$ ) a set of  $Ag_1$ 's *competence beliefs* ( $B-Com_{Ag_1,Ag_2,t}$ ) on  $Ag_2$  about  $\tau$ ;
  - $\Omega iii$ ) a set of  $Ag_1$ 's *disposition beliefs* ( $B-Dis_{Ag_1,Ag_2,t}$ ) on  $Ag_2$  about  $\tau$ ;
  - $\Omega iv$ ) a set of  $Ag_1$ 's *practical opportunities beliefs* ( $B-PrOp_{Ag_1,Ag_2,t}$ ) on  $Ag_2$  about  $\tau$ .
- $\Sigma$ ) A set of mental states ( $MS-REL_{Ag_1,Ag_2,t}$ )—called *reliance*—that must be added to the “core trust” ones and that are strictly linked with the decision to trust; in particular:
  - $\Sigma i$ ) a set of  $Ag_1$ 's *dependence beliefs* ( $B-Dep_{Ag_1,Ag_2,t}$ ) (it is needed or it is better to delegate than not delegate to  $Ag_2$  [6])
  - $\Sigma ii$ ) a set of  $Ag_1$ 's *preference beliefs* ( $B-Pre_{Ag_1,Ag_2,t}$ ) to delegate to  $Ag_2$  (in fact, although this notion is related to the dependence notion, we like to mark it).

We can imagine that each of the beliefs included in  $Trust-Attitudes(Ag_1 Ag_2 \tau t)$  will have a specific weight: a degree of credibility ( $DoC(B_i)$ ), with  $0 \leq DoC(B_i) \leq 1$ .

We consider here the resulting degree of  $Ag_1$ 's trust on  $Ag_2$  about  $\tau$  at time  $t$ , as the simple composition of all these factors.<sup>14</sup>

If we call *Eval-DoT* the function that when applied to a set of mental states returns as result the composition of the weights of these mental states we can have

$$\begin{aligned} & Eval-DoT(Trust-Attitudes(Ag_1 Ag_2 \tau t)) \\ &= DoT_{Ag_1,Ag_2,\tau,t} \quad (0 \leq DoT_{Ag_1,Ag_2,\tau,t} \leq 1). \end{aligned}$$

In order that  $Ag_1$  trusts  $Ag_2$  about  $\tau$  at  $t$ , and then it delegates that task, it is not only necessary that the  $DoT_{Ag_1,Ag_2,\tau,t}$  exceeds

<sup>14</sup>We are not considering in this paper the relevant problem of how this composition should be made: it might be very interesting to analyze the saturation effect (see [5]) of each factor, the possible interference among them, and so on.

a given ( $Ag_1$ 's) threshold, but also that it constitutes the better solution (compared with other possible solutions). So we should consider the abstract scenario<sup>15 16</sup> of Fig. 4.

According to this scenario the following possible choices may be given:

- i)  $Ag_1$  tries to achieve the goal by itself;
- ii)  $Ag_1$  delegates the achievement of the goal to another agent ( $Ag_2, \dots, Ag_n$ );
- iii)  $Ag_1$  does nothing (relatively to this goal), i.e., renounces it.

It is possible to determine a trust choice starting from each combination of credibility degrees— $\{DoT_{Ag_1,Ag_i,\tau,t}\}$  with  $Ag_i \in \{Ag_1, \dots, Ag_n\}$ —of the main beliefs included in *trust-attitudes* ( $Ag_1 Ag_i \tau t$ ) (with  $Ag_i \in \{Ag_1, \dots, Ag_n\}$ ), and from a set of  $Ag_1$ 's utilities  $\{U_{p,t}^+, U_{p,t}^-, U_{di,t}^+, U_{di,t}^-, U_{o,t}\} = U(Ag_1,t)$ , with  $i \in \{2, \dots, n\}$ .

It is possible that—once fixed the set of utilities and the kind and degree of control—different combinations of credibility degrees of the main beliefs produce the same choice. However, more in general, changing the credibility degree of some beliefs should change the final choice about the delegation (and the same holds for the utilities and for the control).

So, if we suppose to have at time  $t_0$  a set constituted by:  $\{DoT_{Ag_1,Ag_i,\tau,t_0}\}$  and  $U(Ag_1,t_0)$ , we will have, as a consequence, the delegation  $Kind_0Delegates(Ag_1 Ag_i \tau_0)$  with  $Ag_i \in \{Ag_2, \dots, Ag_n\}$ , and  $Kind_0Delegates \in \{Performs, Weak-Delegates, Mild-Delegates, Strong-Delegates, Nothing\}$ .

At a different time  $t_1$  we might have a new set  $\{DoT_{Ag_1,Ag_i,\tau,t_1}\}$  and/or a new set of utilities, with

$$\begin{aligned} & \left( \{DoT_{Ag_1,Ag_i,\tau,t_0}\} \neq \{DoT_{Ag_1,Ag_i,\tau,t_1}\} \right) \\ & \quad \cup (U(Ag_1,t) \neq U(Ag_1,t_1)). \end{aligned}$$

So the new delegation could be  $Kind_1Delegates(Ag_1 Ag_i \tau_1)$  with possibly both ( $Kind_1Delegates \neq Kind_0Delegates$ ) and ( $\tau_1 \neq \tau_0$ ).

In the previous section we have seen how, in order to adjust a given delegation/adoption, it is necessary that the agent has specific reasons, that is, new beliefs and goals as compared to the moment of the delegation (adoption).

What, in fact, this means is simply the following.

- The delegator's mental state has changed in at least one of its components in such a way that the action to choose is different from the previous one.

<sup>15</sup> $U(Ag_1)_t$ , is  $Ag_1$ 's utility function at time  $t$ , and specifically:  $U(Ag_1)_{p,t}^+$ , the utility of  $Ag_1$ 's success performance;  $U(Ag_1)_{p,t}^-$ , the utility of  $Ag_1$ 's failure performance;  $U(Ag_1)_{di,t}^+$ , the utility of a successful delegation to the agent  $i$  (the utility due to the success of the delegated action);  $U(Ag_1)_{di,t}^-$ , the utility of a failure delegation to the agent  $i$  (the utility due to the failure of the delegated action);  $U(Ag_1)_{o,t}$ , the utility of doing nothing.

<sup>16</sup>More precisely, we have

$$\begin{aligned} U(Ag_1)_{p,t}^+ &= Value(g) + Cost[Performance(Ag_1 t)] \\ U(Ag_1)_{p,t}^- &= Cost[Performance(Ag_1 t)] + ADF \\ U(Ag_1)_{di,t}^+ &= Value(g) + Cost[Delegation(Ag_1 Ag_i \tau t)] \\ U(Ag_1)_{di,t}^- &= Cost[Delegation(Ag_1 Ag_i \tau t)] + ADF \end{aligned}$$

where ADF = additional damage for failure, and it is supposed that it is possible to attribute a quantitative value (importance) to the goals and where the costs of the actions (delegation and performance) are supposed to be negative.

- The delegee's level of self-trust or the delegee's trust in the environment has changed, and there is some disagreement with  $Ag_i$  about this.

At the same time the new sets of beliefs and utilities might suggest various possible strategies of recovery of the trust situation, i.e., given  $Kind_0Delegates(Ag_1, Ag_i, \tau)$ ,  $DoT_{Ag_1, Ag_i, \tau, t_0}$  and  $U(Ag_1, t_0)$  we might have an adjustment of  $Kind_0Delegates$ .

This adjustment reflects a modification in the mental ingredients. More precisely, the client either *updates* or *revises* its delegation beliefs and goals, i.e.,

- a) either it revises its *core trust beliefs* about  $Ag_2$  (the latter's goals, capabilities, opportunities, willingness);
- b) or it revises its *reliance beliefs* about: i) its dependence on  $Ag_2$ , or ii) its preference to delegate to  $Ag_2$  than to do by itself, or to delegate to  $Ag_3$  (a third agent) or to renounce the goal;
- c) or it changes its risk *policy* and is more or less likely to accept the estimated risk (this means that the client changes either its set of utilities ( $U(Ag_1, t_0)$ ) or its set of thresholds.

In other words, either  $Ag_1$ 's trust on  $Ag_2$  is the same but its preferences have changed (including its attitude toward risk), or  $Ag_1$  has changed its evaluations and predictions about relying on  $Ag_2$ .

Another important role is played by the control that can allow delegation also to a not very trusted agent; for an analysis on this relationship see [19].

## VII. CONCLUSION

Autonomy is very useful in collaboration and even necessary in several cases but it is also risky—because of misunderstandings, disagreements and conflicts, mistakes, private utility, etc. The utility and the risk of having an autonomous collaborator can be object of a tradeoff by maintaining a high level of interactivity *during* the collaboration, by providing *both* the man/delegator/client and the machine/delegee/contractor with the possibility of having initiative in interaction and help (*mixed initiative*) and of *adjusting* the kind/level of delegation and help, and the degree of autonomy run-time. This also means providing channels and protocols—on the delegator's side—for *monitoring* (reporting, observing, inspecting), and for *intervention* (instructions, guidance, helps, repair, brake); and—on the delegee's side—providing some room for discretion and practical innovation; for both client and contractor, channels and protocols are needed for communication and *renegotiation* during the role-playing and the task execution.

Our model provides a principled framework for adjusting autonomy. In particular, we have shown that in order to adjust autonomy one should in fact adjust the delegation/help relationship. Thus, a precise characterization of different dimensions of delegation and of goal-adoption is necessary. Moreover, we argued that adjustment is *bidirectional* (one can expand or reduce the delegee's autonomy) and is *bilateral*; not only the human user or the delegator but also an adaptive/intelligent delegee, the machine (the "agent") can change or try to change its level of autonomy by modifying the received delegation or the previous

level/kind of help. This initiative is an additional and important aspect of its autonomy. We showed how trust, being the mental ground and counterpart of delegation, plays a major role in adjustment: limiting autonomy is usually due to a trust crisis, while expanding autonomy is usually due to an increased trust. Collaborative conflicts are mainly due to some disagreement about the agent's trustworthiness.

We assume that this theoretical framework can be useful also for developing principled systems.

We have outlined

- the criteria about when and why to adjust the autonomy of an agent (for example, when one believes that the agent is not doing (in time) what it has delegated to do and/or is working badly and makes mistakes; and/or one believes that there are unforeseen events, external dangers, and obstacles that perhaps the agent is not able to deal with);
- possible protocols of both monitoring and inspection, and of physical or communicative intervention, that are necessary for control and adjustment.

A very important dimension of such an interaction has been neglected: the normative dimension of empowerment and autonomy (entitlement, permission, prohibition, etc.) which is related with a more rich and institutional relation of delegation [12]. Also this dimension is a matter of run-time adjustment and must be included as a necessary component when modeling several forms of interactions and organizations.

Another important issue for future works is the acceptable limits of the agent's initiative in helping. Would, for example, our personal assistant be too intrusive by taking care of our "interests" and "needs" beyond and even against our request (*hypercritical help*)? Will the user/client like such level of autonomy or would he/she prefer an obedient slave without initiative? Let us leave this question unanswered being enough to have characterized and delimited the complex framework of such an issue.

Finally, we leave to future work a rather important clarification for engineering: does the implementation of such a model necessarily require deliberative agents?

In fact our framework for collaboration and adjustable autonomy is presented in terms of cognitive agents, i.e., of agents that have propositional attitudes, reason about plans, solve problems, and even assume an "intentional stance" by having a representation of the other agent in terms of mind. This can be exemplified via some kind of BDI agent, but in fact it is more general (it does not only apply to a specific kind of architecture). We present our framework in a cognitive perspective because we want to cover the higher levels of autonomy<sup>17</sup>, and also the interaction between a human user and a robot or a software agent, or between humans. However, the basic ontology and claims of the model could be also applied to noncognitive, merely rule-based agents.

<sup>17</sup>In our view, to neglect or reduce the mental characterization of delegation (allocation of tasks) and adoption (to help another agent to achieve its own goals) means, on the one hand, lose a set of possible interesting kinds and levels of reliance and help and, on the other hand, do not completely satisfy the needs and the nature of human interaction that is strongly based on these categories of cooperation.

Obviously, a cognitive agent (say a human) can delegate in a weak or mild sense a merely rule-based entity. Strong delegation based on mutual understanding and agreement cannot be used, but it can be emulated. The delegated device could have interaction protocols and reactive rules such that if the user (or another agent) asks to do something—given certain conditions—it will do that action. This is the procedural emulation of a true “goal adoption.”

Our notions could in fact be just embedded by the designer in the rules and protocols of those agents, making their behavior functionally corresponding to delegation or adoption, without the “mental” (internal and explicit) goal of delegating or of helping.

One could for example have fixed rules of over-help like the following:

⟨if  $Ag_x$  asks for departure time and provides departure time and gate⟩ (over-answering);  
 ⟨if  $Ag_x$  asks for action  $\alpha$  that has result  $r$  and is not able to perform  $\alpha$  but is able to perform  $\alpha'$  with the same result  $r$ , then perform  $\alpha'$ ⟩.

The previous behaviors are in fact a kind of *over-help* although the performing agent does not conceive any help (the real adopter is the programmer writing such a rule).

The same holds in the case of a rule-based delegated agent: the agent could have simple *rules* for abstaining from doing by itself  $\alpha'$  while inducing—via some protocols—the needed action in another agent, or for abstaining from doing by itself  $\alpha'$  when receiving information (communication protocol; observation) about another agent already doing the needed action.

In sum, several basic phenomena and issues (of delegating, adopting, monitoring, intervening, changing delegation, etc.) hold also with and among noncognitive agents and can be incorporated in a procedural emulation of a really social interaction.

#### APPENDIX SCENARIO

In order to exemplify our numerous distinctions, let us imagine a scenario where there are

- an agent E (*explorer*) in charge of moving around in a unknown territory L (*land*) (forest, desert, etc.) exploring and searching for something (T, a “*treasure*”);
- the *Master* (M) of this agent that assigned it such a mission and sent it to L; M is not there (in L) but he follows E at a distance (either he can observe E and some part of L, or he is in contact with E via some communication media);
- an enemy, a *spy* S who is observing and following E, in order to catch its secrets and gain T.

This is an imaginary scenario, but its characteristics can be applied to an autonomous robot exploring space, or to a software agent trying to do something for the user, or to a smart house with many autonomous objects, and so on.

*Ex1 (Weak delegation or reliance):* S is observing E during its search and relies upon E for such an exploration and for finding T. E is not aware of S observing and exploiting it.

*Ex2:* (In the same situation as Ex1): Now E is aware of S observing and exploiting it but E does not worry about this because E is sure to control the situation.

*Ex3:* (In the same situation as Ex1): Now E is aware of S observing and exploiting it and he is worried about this. So, E changes his behavior (for example, he pretends to achieve goals that he does not have, or masks some actions, and so on).

*Ex4:* E is now aware of S observing and exploiting it. This belief introduces a new goal in E’s mind, for example, to deceive S about where T is, or to kill S, and so on.

If S does not know about this it will continue to try to exploit E.

*Ex5 (Mild delegation):* S has introduced in E’s compass a microchip that changes the direction of the needle. S uses this trick each time he risks to be seen by E. E’s behavior is influenced by this trick.

*Ex6 (Strong delegation):* E works for M; he knows and agrees upon exploring L and sending data and results to M, who relies on E for exploring L and finding T.

*Ex7 (Spontaneous unilateral help):* E believes that—in a specific area—M is observing him via a camera; he understands (without any previous agreement) that if he is in dark areas M cannot see him. He decides—without being requested or even without M realizing this—to work in visible areas.

Or S realizes that there is a serious danger (say, a predator) for E, and decides to terminate it. E is unaware of this (or E realizes it afterwards).

*Ex8 (Strong help):* E believes that—in a specific area—M is observing him via a camera; he understands that if he is in dark areas M cannot see him. He offers to work in visible areas on condition that M increases E’s reward (in fact, to work in these new conditions is more dangerous than in the dark area).

Or S realizes that there is a serious danger (say, a predator) for E, and offers its help to E in exchange for its trust.

*Ex9 (Close delegation):* M says to E: “There is a building in front of you, right? Go straight ahead, enter the main door, go upstairs, first corridor, third door on the left.”

*Ex10 (Open delegation):* In the same situation as before: either M says to E: “There is a building in front of you, right? You should arrive at the third room, first corridor, first floor.” (“To arrive” is an abstract action; there are several ways for arriving and several kinds of arriving).

Or M says to E: “There is a building in front of you, right? You have just five minutes to be in the third room, first corridor, first floor. Find some way to get in.” (“To be in” is a desired state of the world; no action is specified).

*Ex11 (From weak to mild delegation—initiative of delegator):* S was relying on E for exploring a given area (weak delegation and E ignores the existence of S. Now S realizes that in fact E is going in a different direction. S could provoke a landslide to influence E to change his plan: E’s going in the right direction (and S exploiting this) is now moved by S’s influence. This is an example of mild delegation.

*Ex12 (From weak to strong delegation—initiative of delegator):* Suppose that there is another agent H (hunter) exploiting E’s action of exploring a given area to drive out some prey. He relies on E for systematically exploring a part of L. Now H observes that E didn’t examine a part of the current

zone, and he explains his goal to E and asks him to also explore this part in exchange for some money. In such a way H attempts to establish a strong delegation based on his explicit request and reliance and on E's awareness and agreement.

*Ex13 (From weak to strong delegation—initiative of delegee):* In the same situation of Ex12; this time it is E who is aware of the fact that H is exploiting his actions: so, E asks H to pay him in exchange for his help.

*Ex14 (From weak to strong adoption—initiative of adopted agent):* S is getting some predators out of E's way and E is aware of S's help (even if he does not know the final goal of S). E asks S to find an agreement about this help because, for example, he wants to be sure that S will continue to help him in the future or because he wants to better explain how that help must be given.

*Ex15 (From weak to strong adoption—initiative of adopting agent):* S is getting some predators out of E's way. S can look for an agreement about this help because he wants to be sure that E's behavior will be coherent with this help.

*Ex16 (Reduction of help):* M says to E: "Open that door and take the statue through it, " but E limits himself to opening the door and going through it because he is not able to raise the statue.

*Ex17 (Extension of help):* M to E: "Open that book and read what is written there." E knows that M can observe him via a camera and while reading the book he shows it to the camera so that M can read and check.

*Ex18 (Critical overhelp):* M says to E: "Open that book and read what is written from page 20 to 30." E knows that M wants to know the instructions for making a chemical explosive, and he extracts the explosive directly from a bag and says: "What should we explode?"

#### REFERENCES

- [1] G. Ferguson and J. Allen, "TRIPS: An integrated intelligent problem-solving assistant," in *Proc. Nat. Conf. AI*, Menlo Park, CA, 1998.
- [2] M. Hearst, "Mixed-initiative interaction—Trends & controversies," *IEEE Intell. Syst.*, vol. 14, pp. 14–23, Sept./Oct. 1999.
- [3] H. Hexmoor, Ed., *J. Experimental Theoretical Artif. Intell., Special Issue Autonomy Contr. Software*, vol. 12, no. 2, Apr.-June 2000.
- [4] C. Castelfranchi and R. Falcone, "Toward a theory of delegation for agent-based systems robotics and autonomous systems," *Robot. Autonom. Syst., Special Issue Multi-Agent Rationality*, vol. 24, pp. 141–157, 1998.
- [5] —, "Principles of trust for MAS: Cognitive anatomy, social importance, and quantification," in *Proc. Int. Conf. Multi-Agent Syst*, Paris, France, July 1998, pp. 72–79.
- [6] J. Sichman *et al.*, "A social reasoning mechanism based on dependence networks," presented at the 11th ECAI Conf., 1994.
- [7] M. Pollack, "Plans as complex mental attitudes," in *Intentions in Communication*, P. R. Cohen *et al.*, Eds. Cambridge, MA: MIT Press, 1990, pp. 77–103.

- [8] C. E. Martin and K. S. Barber, "Multiple, simultaneous autonomy levels for agent-based systems," in *Proc. Fourth Int. Conf. Contr., Automat., Robot., Vision*, Westing Stamford, Singapore, 1996, pp. 1318–1322.
- [9] C. Castelfranchi, "Guaranties for autonomy in cognitive agent architecture," in *Intelligent Agents*, M. Wooldridge and N. Jennings, Eds. New York: Springer-Verlag, 1995, vol. LNAI 890, pp. 56–70.
- [10] —, "Founding agent's autonomy on dependence theory," presented at the ECAI Conf., Berlin, Germany, Aug. 2000.
- [11] M. Huhns and M. Singh, *Agents and Multiagent Systems: Themes, Approaches, and Challenges in Reading in Agents*, M. Huhns and M. Singh, Eds. San Mateo, CA: Morgan Kaufmann, 1997.
- [12] A. J. I. Jones and M. Sergot, "A formal characterization of institutionalized power," *J. Interest Group Pure Appl. Logics*, vol. 4, no. 3, pp. 36–52, 1996.
- [13] A. Haddadi, *Communication and Cooperation in Agent Systems*. New York: Springer-Verlag, 1996.
- [14] C. Castelfranchi and R. Falcone, "From task delegation to role delegation," in *AI\*IA: Advances in Artificial Intelligence, Lecture Notes in Artificial Intelligence*, vol. 1321, M. Lenzerini, Ed., 1997, pp. 278–289.
- [15] P. Cohen and H. Levesque, "Rational interaction as the basis for communication," Center Study Lang. Inform. (CSLI), Stanford, CA, Tech. Rep., 1987.
- [16] N. R. Jennings, "Commitments and conventions: The foundation of coordination in multi-agent systems," *Knowledge Eng. Rev.*, vol. 3, pp. 223–250, 1993.
- [17] C. Castelfranchi and R. Falcone, "Conflicts within and for collaboration," in *Conflicting Agents: Conflict Management in Multi Agent Systems*, C. Tessier *et al.*, Eds. New York: Kluwer, 2000, pp. 33–61.
- [18] C. Castelfranchi, "Commitment: From intentions to groups and organizations," presented at the ICMAS Conf., San Francisco, CA, June 1996.
- [19] C. Castelfranchi and R. Falcone, "Trust and control: A dialectic link," *Appl. Artif. Intell. J., Special Issue on Trust in Agents*, pt. 1, vol. 14, no. 8, 2000.



theory.

**Rino Falcone** received the Laurea degree in physics from the University of Rome, Rome, Italy, in 1984.

He is currently a Computer Scientist with a background in physics. He has worked in the Voice Laboratory Research Department of Olivetti, Torino, Italy, and is now a researcher for the Division of Artificial Intelligence, Cognitive, and Interaction Modeling, Institute of Psychology, Italian National Research Council (CNR), Rome. His main scientific competence and experience ranges from natural language processing to multiagent systems and agent



**Cristiano Castelfranchi** received the Laurea degree in physics from the University of Rome, Rome, Italy, in 1969.

He is a Former Senior Researcher with the Italian National Research Council (CNR), Rome. His main background is in linguistics and cognitive science. He has published mainly in the area of artificial intelligence (AI), distributed AI, and multi-agent systems. Currently, he is a Full Professor in General Psychology, Department of Communication Sciences, University of Siena, Siena, Italy.