Commitments: From Individual Intentions to Groups and Organizations

Cristiano Castelfranchi

PSCS - Project for the Simulation of Social Behaviour
Istituto di Psicologia/CNR - Roma, Italy
cris @pscs2. irmkant. rm. cnr. it

Abstract

The aim of this work is to introduce some notions of Commitment as a descriptive ontology crucial for the understanding of groups' and organizations' functioning, and of the relations between individual agents and collective activity. Some of the basic ingredients of such notions are identified and some steps are made towards their definition. In particular, it is claimed that a notion of Commitment is needed as a mediation between the individual and the collective one. Before constructing a notion of "Collective or Group Commitment" a notion of "Social Commitment" is to be defined. "Social commitment" is not an individual Commitment shared by many agents; it is the Commitment of one agent to another. The normative contents (entitlements / obligations) of this social relation are stressed and its connections with individual intentions and collective activity are alalized. On that basis, a notion of Organizational Commitment is proposed, that could account for the structure of stable Organizations. Commitment is a crucial notion both to analyse the structure of Organizations and to support cooperative work, but a deeper analysis is needed, connecting agent's mental states with social relations and structure.

Introductive remarks

There is an implicit agreement about Organizations in recent computational studies. Either in DAI theories of organization [Bond 1989] [Gasser 1991], or in formal theories of collective activity, team or group work, joint intention, and "social agents" [Levesque, Cohen, & Nunes 1990] [Rao, Georgeff, & Sonenberg 1992], or in CSCW approaches to cooperation [Winograd 1987], *Organization is accounted for by means of the crucial notion of "commitment"*. "Commitment" is seen as the glue of the group, of collective activity: it links the agent with the joint goal and the common solution, it links members' actions with the collective plan, it links the members with each other.

Unfortunately, the current analysis of Organizations in terms of Commitment is quite unsatisfactory, for a number of reasons:

a) the current definitions of Commitment are insufficient to really account for stable group constitution and activity;

b) there is a dangerous confusion between the notion of "social" and that of "collective", so there is no theory of "social" commitment as a necessary premise for a theory of collective or group commitment;

c) the relationships among the personal Commitment to an action (implied in the notions of "intention" and "intentional act"), the group's Commitment to the collective act, and the Commitment of a member to his group and the collective activity, are not clearly stated;

d) agents seem to be completely free (also in Organizations) to negotiate and establish any sort of Commitment with any partner, without any constraint of dependence and power relations, of norms and procedures, of pre-established plans and cooperations [Castelfranchi, Conte, & Cesta 1992].

The present work discusses these problems which are crucial to the assessment of the *necessary ontology* for describing and understanding cooperation among autonomous agents, group activity, and Organizations. This is not a formal paper, we just claim that the clarification of such an ontology is a preliminary step toward a formal definition of cooperative relations, and that it is also necessary for improving the exchange between AI approaches to multi-agent systems, and social and management approaches to cooperative work.

In particular, we will analyse a notion of "Social Commitment" as distinct on one side from the notion of "Individual-Internal commitment", and from the notion of "Collective Commitment", on the other side. This notion of *Social Commitment is a relational one*; it cannot be defined with regard to one single agent (be it an individual or a group). However, this notion is not purely behaviourist, like in [Singh 1992]: *it could be analysed in terms of the mental states of the partners*. We will also try to show that it is not reducible to the individual Commitments of the partners, and under which conditions it implies such individual Commitments.

On such a basis (and on the basis of the notion of Generic and of Meta-Commitment) we would propose: a characterisation of the kind of Commitment that supports the structure of an Organization, including the notion of

role", the claim that there is *no Organization without Obligations.*

This approach is necessary for a good definition of our concepts and for developing a formal theory of groups, organizations, collective actions. Without such a kind of analysis it seems impossible to account for different kinds of organization (e. g. strictly cooperative Vs orchestrated), and consequently for different kinds of Commitment (by role or by reciprocation, free, based on interest or benevolence, etc.), and then for different ways to solve conflicts, both intra- and inter- agents. For ex. how to intervene when one of the agents decides to abandon the group, or the common plan? We cannot influence appropriately this agent if we don't know exactly his kind of Commitment and the related different reasons for defection [Castelfranchi & Cesta 1993].

Kinds of Commitment: Internal, Social, and Collective

Before proposing a distinction between "social" and "collective", it is necessary to go back to the notion of individual (or better "internal") commitment.

INTERNAL Commitment (I-Commitment)
Internal Commitment --as [Boudon 1992] calls it-- corresponds to the Commitment defined by Cohen and Levesque (on the basis of Bratman's analysis) [Cohen & Levesque 1990]. It refers to *a relation between an agent and an action.* The agent has decided to do something, the agent is determined to execute a certain action (at the scheduled time), the goal (intention) is a persistent one. The way to capture such a persistence is to establish that the intention will be abandoned only if and when the agent believes that the goal has been reached, or that it is impossible to achieve, or that it is no longer motivated. We will refer to the formal notions of belief, goal, knowledge, defined in that approach [Cohen & Levesque 1990] [Levesque, Cohen, & Nunes 1990].

The term "Internal" is to be preferred to "Individual" because one may attribute I-Commitments also to a group. Thus, an "individual Commitment" is just the I-Commitment of an individual agent. The term "psychological", as opposed to "social" [Singh 1992], is quite misleading because also the Social and the Collective Commitments are relations among minds.

SOCIAL Commitment (S-Commitment)
We need a notion of Commitment as a mediation between the individual and the collective one. *A "social commitment" is not an individual Commitment shared by many agents.*

In general, it is absolutely necessary to distinguish between the notion of "social" and the notion of "collective". This distinction is not addressed in DAI and MAS, where "social" is already "collective", for ex. to characterize a goal or intention shared by many agents and unachievable independent of each other, it is used the notion of "social" intention or goal (ex. [Werner 1988]); to designate an agent formed by many individual agents, i. e. a group or a team, it is used the notion of "social agent" (ex. [Rao, Georgeff, & Sonenberg 1992]); to denote the idea of reciprocal commitments within a team of agents it is used the notion of "social Commitment" and "social plan". *"Social" is not a synonym of "collective".* There is a very important level of "social action", "social agent", and "social mind", where action and mind remain "individual" while in the meantime they are oriented toward another social entity [Conte & Castelfranchi 1995]. At this level, before constructing a notion of "collective or group Commitment" we need a notion of "social Commitment": the Commitment of one agent to another [Singh 1992].

As said above S-Commitment is a relational concept. *It expresses a relation between at least two agents.* More precisely, S-Commitment is a 4 argument relation:

(S-COMM x y a z)

where x is the committed agent; a is the action x is committed to do; y is the other agent to whom x is committed; z is a third agent before whom x is committed.

Let us neglect (in this paper) the third agent (z), i. e. *the witness*, who has a very crucial role in normative contexts (norms efficacy) [Conte & Castelfranchi 1993] and in contractual contexts implying also free riders and cheaters. Here, we will focus on the relation between x and y.

COLLECTIVE Commitment (C-Commitment)
We should differentiate S-Commitment from the Collective Commitment or Group Commitment. The latter is just *an Internal Commitment of a Collective agent* or Group. In other terms, a set of agents is Internally Committed to a certain intention and (usually) there is mutual knowledge about that. It remains to be clarified: a) if such conditions are sufficient to account for the Commitment to a collective activity; b) which are the relationships between S-Commitment and C-Commitment (later).

Toward a Definition of Social Commitment

One might say that S-Commitment is just an I-Commitment of x to a known by y. S-Commitment is much more than that. In fact, this condition is insufficient. Let us suppose that y is aware of some intention of x to do an action a: this does not allow to say that "x is committed to y to do a". Nor is it sufficient to suppose that x is aware that y knows his intention, etc. (Mutual Knowledge). Let us give an example: we have realised that John intends to play a bad joke on Paul, and John becomes aware of the fact that we understood his intention; nevertheless, John persists in his plan, relying on our silence. We may keep silent and become party to his trick. Even in this case it would be false that "John is committed to us to play a joke to Paul". If John drops out

his I-Commitment (that we know), we can be surprised, but we are not entitled to protest. In this situation two qualifying aspects of the social relation of "being committed to somebody" are lacking:

a) *A social Commitment is a form of "Goal Adoption".* In other terms: x is committed to y to do a, if y is interested in a. The result of a is a goal of y;for this reason, y has the goal that x does a. Thus we should include in the formal definition of S-Commitment the fact that *(S-COMM x y a z)* implies that *(GOAL y (DOES x a))*.

So, normally (see 7th §) *x and y share a goal: (DOES x a)*. Notice that:

- x and y have an identical goal, but in virtue of x 's adoption of y 's goal;
- this goal for x is an Intention (since it is his own action);

b) *If x is S-Committed to y, then y can (is entitled to):*

- **control** if x does what he "promised";
- **exact/require** that he does it;
- **complain/protest** with x if he doesn't do a;
- (in some cases) **make good his losses** (pledges, compensations, retaliations,.)

We will come back later to this necessary and very relevant ingredient of the S-Commitment: the rights of y on x created by the S-Commitment of x to y.

At this point, necessary elements for a formal definition of S-Commitment appear to be the following ones:

x and y mutually know that x intends to do a and that this is y 's goal, and that as for a y has specific **rights on** x *(y* is entitled by x to a*)*.

One should introduce a relation of "entitlement" between x and y meaning that y *has the rights* of controlling a, of exacting a, of protesting (and punishing), in other words, x *is S-Committed to y to not oppose to these rights* of y (in such a way, x "acknowledges" these rights of y). So, interestingly enough, the definition of S-Commitment seems to require a recursive call of another S-Commitment of x itself (see 6th §).

Other Relevant Aspects of Social Commitment

Five main aspects deserve to be mentioned:

a) *Not all the adoptions of a goal of y by x imply a S-Commitment of x to y.* What else is required? First, the Mutual Knowledge we already mentioned. Second, *y 's agreement* !

In fact, if x has just the I-Commitment to favour one of y 's goals, this is not sufficient (even if there is common awareness): y should "accept" this. In other words, she decided, she is I-Committed to achieve her goal by means of x's action. This acceptance is known by x, there is an agreement. Then, *the S-Commitment of x implies a S-Commitment of y to x to accept x's action* (y doesn't refuse, doesn't protest, doesn't say "who told y ou!", ...). Without such (often implicit) agreement (which is a reciprocal S-Commitment) no true S-Commitment of x to y has been established.

b) What stated just now does not implies necessarily explicit communication between x and y; we mean communication by apposite message sending. Implicit Commitments (with implicit agreements) are also allowed. One can characterize the *principle of implicit committing* as follows:

IF

there is mutual knowledge between x and y about an *expectation* of y about an action of x (where an *expectation* is a *believe* about a future state or action plus a *goal* about the same state or action);

and

x intends to do that action also because he knows about y 's expectation;

and

x does not explicitly deny his intention, does not contradict y 's expectation;

THEN

x implicitly takes a S-Commitment to y for that action;

y is entitled by default to consider x S-Committed to her.

One could consider this "understanding" among agents as a form of "implicit communication" (communication without specialized message sending). On this perspective one could also say that S-Commitments always require communication.

c) *The very act of committing oneself to someone else is a "rights-producing" act*: before the S-Commitment, before the "promise" [Searle 1969], y has no rights over x, y is not entitled (by x) to exact this action. After the S-Commitment it exists such a new and crucial social relation: y has some rights on x, she is entitled by the very act of Commitment on x's part. So, the notion of S-Commitment is well defined only if it implies these other relations:

- y is entitled (to control, to exact/require, to complain/protest);
- x is in debt to y ;
- x acknowledges to be in debt to y and y 's rights.

In other terms, x cannot protest (or better *he is committed to not protesting*) if y protests (exacts, etc.).

d) What we said just now implies also that *if x is S-Committed to y, he has a duty, an obligation*, he ought to do what he is Committed to:

(S-COMM x y z a) $\supset$ *(OUGHT x (DOES x a))*.

So, when x is committed, a is more than an Intention of x, it is a special kind of goal, more cogent. For a complete theory of S-Commitment, of its representation and of its social effects, we need also a theory of the mental representation of obligations, a definition of a predicate to express such an obligation [Conte & Castelfranchi 1993.], and a theory of the different sources and kinds of goals that cannot be considered "desires" or "wishes" (like duties)!

y's rights and *x*'s obligations could give the impression of a very unbalanced social relation. However, one should consider that normally in cooperation, in social exchange, in contracts, in Organizations S-Commitments are Reciprocal. For example, a worker is committed to his boss, and his boss is committed to him: they both have rights and duties.

e) The more cogent and normative nature of S-Commitment explains why abandoning a Joint Intention or plan, a coalition or a team is not so simple as dropping a private Intention. This is not because the dropping agent must inform her partners [Levesque, Cohen, & Nunes 1990] [Jennings 1993] - behaviour that sometimes is even irrational [Castelfranchi & Cesta 1993]-, but precisely because Joint Intentions, team work, coalitions (and what we will call Collective-Commitments) imply S-Commitments among the members and between the member and her group (8th §). In fact, one cannot exit a S-Commitment in the same way one can exit an I-Commitment. Consequences (and thus utilities taken into account in the decision) are quite different because in exiting S-Commitments one violates obligations, frustrate expectations and rights she created. We could not trust in teams and coalitions and cooperate with each others if the stability of reciprocal and collective Commitments was just like the stability of I-Commitments (Intentions).

Such creation of interpersonal obligations and rights through S-Commitments ('micro-deontics') will require a general approach to deontics that allows contradictions among deontic contexts and hierarchical levels (in this direction, see e.g. [Jones & Porn 1985]). For example, a killer gets an obligation to his instigator to murder somebody, but, from the point of view of the society such an obligation is in contrast with a prohibition (law) and with a much stronger obligation.

S-Commitment satisfaction

For a correct ontology about commitment relationships in group and organizations it is also necessary to distinguish between *Subjective Commitment satisfaction* and *Objective Commitment satisfaction.*

In his proposal for a Commitment-based analysis of cooperative work, [Fikes 1982] suggested that the basic criterion for successful task completion and Commitment fulfilment was the *"satisfaction of the client"*. In our view, this criterion follows (or anticipates) the common subjective bias of AI and Cognitive Science [Conte, Miceli, & Castelfranchi 1991], and it is very fallacious and dangerous.

In our analysis, what the "client" *y* asked was: (a) *the satisfaction of her goal/need.*

She did not ask the "contractor" *x*: (b) *to make her believe that her goal is achieved.*

This is quite different! The goal "to know that her goal is achieved" is for the "client" *y*, just a "control goal" of her main and substantive goal [Castelfranchi 1992]. If the "contractor" *x* achieves only the condition (b), the "client" will be satisfied (until she ignores the truth). However, the "contractor" *x didn't satisfy the Commitment* and the implied Obligation. In fact, if there is a bystander (the "witness" *z*), he can put blame on *x* (and *x* may not protest for this); if the "client" *y* discovers the truth she will be entitled (by the original promise) to claim and protest (she is now even more entitled by *y* 's defeat and lie).

This problem is quite relevant also in applications. In current CSCW systems for example the "performance" phase (the execution of the promised action) in fact is just a communicative act: a declaration by the contractor.

Condition (b) (*Subjective satisfaction*) is neither sufficient nor necessary for true Commitment satisfaction. It is *not sufficient*, as we saw, because it exposes the "client" *y* to fraud, and because all rights and obligations of the Commitment remain valid (and unsatisfied). It is *not strictly necessary*, because in many cases agents make S-Commitments to other agents when they mutually know that it is not the case that *y* will be informed about the fulfilment. This is the case of promises to dying people (for ex. testaments).

So, when is the task completed, when is the Commitment satisfied? It is true that normally condition (a) (*Objective satisfaction*) is insufficient. Without *y* 's being aware of Commitment fulfilment, expectations and claims remain suspended within *y* 's mind. *y* has the goal to know whether *x* completed his task; and *x* has the goal that *y* knows that all "debts" are extinguished. In some sense, normally (by default and implicitly) *x is S-Committed to let y know that he completed the task.* They have this joint goal, and there is an implicit agreement about this.

Let us say that: if *x* is S-Committed to *y* to do *a* (and he can inform *y* about his doing *a*), *x is also S-Committed to y to inform y about his doing a.*

Recursive S-Commitment

It seems that an adequate definition of S-Commitment should contains some recursive call to S-Commitments between *x* and *y*:

- the *S-Commitment of x to y to not oppose to y's rights* created by *x*'s committing himself (3th §);
- the *S-Commitment of x to y to inform y* about the expected action (if possible) (5th §);
- the *S-Commitment of y to x to accept x's action* for her goal (goal adoption) (4th §)

Relationships between Social and Internal Commitment: Honesty

S-Commitment is established through a communicative act, be it explicit or implicit. It implies the intention of *x* that "y believes that *x* is S-Committed to *y* to do *a*" and

"that *y* believes *x* is I-Committed to do *a*". This is because in S-Committing himself *x* wants that *y* intends to accept and to control his behaviour (so that *y* 's knowledge is a necessary condition for *x*'s S-Committing). This communicative act has very relevant consequences on the relations between S-Committment and I-Commitment.

S-Commitment may be "sincere" or "insincere", or better, the agent may be "honest" or "dishonest"; when *x* just lets *y* believe that he intends to do what *y* needs. If the agent *x* is "honest", *x*'s S-Commitment to *y* to do *a* also implies an I-Commitment to do *a*. But, if *x* is "dishonest", if the act of socially-commiting himself in "insincere", in this case he is not I-Committed to do *a*, while *he is still S-Committed to do so*, and he actually got an Obligation to do *a*. So:

1) *x's I-Commitment on a is neither a necessary nor a sufficient condition for his S-Commitment on a.* Just *y*'s belief that *x* is I-Committed to *a*, is a necessary condition of *x*'s S-Commitment.

2) we could define an agent as HONEST exactly on this basis:

(HONEST *x*) = def (S-COMM*x y z a*) ⊃ (I-COMM *x a*)

Anyway, let us simplify, postulating that our agents are always "honest" like in other models of Commitment. (That is why there was Mutual Knowledge in our first definition - 3th §). Given this postulate, we may remark that

the S-Commitment of x to y to a implies an I-Commitment of x to a.

In our approach, it is possible to state such a precise relation between the two notions, because also the S-Commitment is analysed in terms of the mental states of the partners. It is not a primitive notion (like in [Singh 1992]). However, it is not "reducible" to the I-Commitment, because it is an intrinsically relational /social notion (among agents), and contains much more than the I-Commitment of the involved agents.

Relationships among Internal, Social, and Collective Commitment

Is a true CollectiveCommitment of a group of agents just the sum of the I-Commitments of the members, or does it require S-Commitments among those members?

We think that there is no univocal answer. It depends on the kind and nature of the group [Conte, Miceli, & Castelfranchi 1991] [Conte & Castelfranchi 1995].

Commitments in strictly "cooperative" groups Let us first consider a true "cooperative" group which in our sense is based on a Common Goal and Mutual Dependence. More precisely, *true cooperation is defined in terms of Mutual Knowledge of Mutual Dependence* relative to an Identical goal of *x* and *y*. Mutual Dependence is defined as follows: *x* Depends on *y* for doing the action *a1* relative to his goal *p*, and *y* Depends on *x* for doing the action *a2* relative to her goal *p*.

An agent *x Depends on an agent y for doing an action a* relative to a state *p* [Castelfranchi, Miceli, & Cesta 1992] [Sichman et al. 1994] when:

- *p* is a goal of *x*'s
- *x* is unable to do any action *a* (and to realise *p*)
- *y* is able to do *a*
- action *a* is in a plan useful to achieve *p*

If each one knows to be dependent on the actions of other agents, each one wants that these others do their share and wants to do his own, for the common goal. Then, in a fully cooperative group, a S-Commitment of everybody to everybody arises: each one has to do his own job. Given that the members form the group, we may say that *each member is S-Committed to the group* to do his share [Singh 1992].

So, the Collective-Commitment (defined as the I-Commitment of a collective agent) will imply (at least in the case of a fully cooperative group):

- the S-Commitment of each member to the group: *x* is S-Committed not simply to another member *y*,but to the all set/group *X* he belongs to;
- the S-Commitment of each member to each other; then also many Reciprocal Commitments;
- the I-Commitment of each member to do his action.

Who is entitled to protest, given that S-Commitment is characterized by entitlements ? The group: each member (or the authority in charge, if such authority exists).

Commitment and the notion of Collective Agent. To define a notion of Collective Agent just on the basis of the I-Commitment (or of the individual intention), of its sharing among the members, and on the members' mutual knowledge, is, in our view, a fallacious attempt.

Members of a "social agent" -- in [Rao, Georgeff, & Sonenberg 1992]'s terms -- or of a group or "team" -- in [Levesque, Cohen, & Nunes 1990]'s -- have a Joint-PersistentGoal, that is, a realizable achievement goal associated with a mutual belief that other members of the team have an equal goal and belief. Now, it is possible to show that this notion is not sufficient to account for a teamwork [Conte & Castelfranchi 1995].

Consider this real life example: the case of two people stranger to each other standing at the same bus stop. It is an optional stop for the driver (i. e. he will stop only if people make some signal to stop). The two persons know that both are waiting for the same bus: so, they have an identical achievement goal ("bus n°3 stops") and they mutually know it. Is this a sufficient basis for a collective activity? Do they form a collective agent, a team? Suppose that unfortunately, each of them in virtue of their equal goal has the expectation that the other does the necessary actions; each of them relies upon the other for achieving the shared goal. The possible result is that the bus arrives, nobody makes the signal, the bus doesn't stop: both of them lose the opportunity to reach the shared goal. What is needed? It is needed an (implicit or explicit) agreement about a common activity, based on the awareness of the reciprocal dependence relations.

Let us give another crucial example. Prof. Montaigner, of the Institute Pasteur in France, and Prof. Gallo in the US both have the final goal *p* "vaccine anti-AIDS be found out" relative to the belief *q* that "if vaccine is found out, AIDS is wiped out". They share all three mental attitudes described [Levesque, Cohen, & Nunes 1990] as necessary and sufficient conditions for a Joint Persistent Goal and then for a team:

1) they mutually believe that *p* is currently false;

2) they mutually know they all want *p* to eventually be true;

3) it is true (and mutually known) that until they come to believe either that p is true, that *p* will never be true, or that *q* is false, they will continue to mutually believe that they each have *p* as a "weak achievement goal" relative to *q* and with respect to the team. Where a "weak achievement goal" with respect to a team has been defined as "a goal that the status of *p* be mutually believed by all the team members".

But no-one would stretch oneself up to saying that Prof. Gallo and Prof. Montaigner form a team. Indeed, given their "parallel goals" ("I find out the vaccine"), they might come to strongly compete with each other.

What else is needed for them to form a team? Without the belief about the mutual dependence, also the Commitment to do one's own share is unmotivated, it is irrational. The belief in an dependence relation or in a "necessity to collaborate" is a basic condition for a really cooperative work and team [Jennings 1993] [Conte, Miceli, & Castelfranchi 1991].

Now, our point is the following: if the members acknowledge their mutual dependence and the need to collaborate, if consequently they (implicitly or explicitly) agree about a (already specified or to be specified) common activity, have they got only an I-Commitment or better have they got S-Commitments to the others? *Is the S-Commitment a necessary condition for the constitution of a Collective Agent?*

For sure, shared or mutual *knowledge* of another's intention is insufficient to account for Joint Intentions. A necessary ingredient of Joint Intention (which is in our notion of S-Commitment) is the *goal* that the partner intends and performs his/her action (task).

Commitments in other kinds of collective activity and group structure. In our view, the current characterization of group activity and collective agent is neither sufficient (as we saw) nor necessary. In many kinds of natural group, team, organization, people participate in a collective activity without sharing the same goals or the ultimate end of the group or organization. Nor are they even expected to have such joint mental states. Let us mention just another "cooperation" model we call *"orchestrated cooperation"* [Conte & Castelfranchi 1995]. Suppose there are three agents: a boss A, and two executors B and C. Suppose that only A knows the end goal of the coordinated activity he requires from B and C, only A knows the complete plan and the reciprocal dependence relations between B and C. It is even possible that B and C ignore each other. The cooperative plan is in one and only one mind: the boss'. Between respectively A and B, and A and C, there are S-Commitments based on social exchange relations (not on true cooperation): B and C are not interested in the result of the plan but just in their personal benefits (rewards). There is no S-Commitment between B and C, who in fact collaborate in a coordinated way, and are members of a team.

Even in these cases we may say that *the group is I-Commmitted to do a* (or to achieve a certain goal), because it is the explicit plan and intention that organizes the action of the participants and determine their S-Commitments and their I-Commintments. But, in this case, *the I-Commitment of the group* (C-Commitment) *doesn't correspond to identical I-Commitments of the members*. However, there is an important constraint on the relation between the members' I-Commitments and the C-Commitment: the former should be "instrumental" to the latter. In other terms, *the S-Commitments of the members (and their consequent I-Commitments) should be Commitments to an action wich is part of the plan (complex action) the group is C-Committed to achieve.*

Generic Commitment, Meta-Commitment and Organizational Commitment

To account for coordination and negotiation among agents in stable groups or in Organizations S-Commitment is insufficient. Something more is needed in order to explain the structure of this long term forms of cooperation, and distinguish between different levels and forms of inter-agent Commitments. In particular, two notions seem to be crucial: Generic Commitment and a Metalevel S-Commitment.

A GenericCommitment is a Commitment to a class of actions: *x* is Committed to do any instance of such a class *A*, any action of that kind:

(GenericCOMM x A) = for all *a* (where *a* is an instance of *A*) *(COMM x a)*

True Organizations are not extemporary, built up at the moment. They are not made of Commitments to do a specific action at a specific moment: so they use Generic-Commitments.

More precisely, they are made of Generic Meta-Commitments: *Commitments to Commit oneself* to do the right thing at the right moment. *These Commitments to Commit oneself determine the "structure" of the Organization*. They are different from the running Commitments involved in the structuring of the collective activities of the Organization. Running Commitments mainly are just instantiations of Meta-Commitment. Thus, *the structure of the Organization is different from the structures of its activities.* The former partially determines the latter.

This notion of Organizational Commitment (Org-Commitment) as Meta and Generic S-Commitments, implies that, for all actions *a* instances of class *A*, if *x* knows that his Organization/Group wants he performs such an action, he has a S-Commitment to his Organization to do *a*.

When the member *x* of the group *X* is organizationally Committed to his group, he is Committed to accept the requests of the group within a certain class of actions (his *office*). Then, *x*'s Org-Commitment to *X* implies that if there is a request of *X* to *x* about an action of the class *A*, *x* is automatically S-Committed to *X* to this action, he automatically gets an obligation to do *a*.

The generic nature of the Org-Commitment may give rise to controversies. In fact, there may be different points of view between *x* and *X* (or *y*) as to whether a specific requested action *a* is or isn't a member of the class *A*.

The **"Role"** of an agent is relative to the group or the Organization. It is *the set of the Org-Commitments of the member to his group*. So, the organizational role is neither a functional behavior, nor a task or a set of tasks [Werner 1988]. It is a "normative" notion: it is a set of behavioral obligations based on the Organizational agent's Commitments to the group relative to certain classes of actions to be requested or expected (tasks).

A Generic S-Commitment of *x* to *y* to *A* (like the Org-Commitment), gives to *y* a special *"power of influencing" x* [Castelfranchi 1990], a **"Command Power"** over *x*. In fact, *x* is committed to comply with any request of *y* as for *A*: *x* is "benevolent" for special reasons towards *y* as for *A*. This prevents *y* from the need to negotiate each time *x*'s compliance with *y* 's requests. This kind of "benevolence" based on such a Commitment-obligation is what we call "obedience".

This "Command Power" gives *y* (or *X*) a very special faculty: that of taking decisions for *x*, of deciding that *x* does something. This *possibility of deciding about other agents' actions* is quite strange: *y* has an "intention" about *x*'s action. More than this, *y can take a S-Commitment for x*, she can Commit *x* to *w* to do something. To be more precise:

- *y* S-Commits herself to *w* about an action of *x*, or better to require and to influence *x* to do this action (*y* is "responsible" for *x*'s action);
- this S-Commitment of *y* to *w* implies an I-Commitment of *y* to require/influence *x* to do *a*;
- then, *y* will require *x* to do *a*. ;
- given that *x* is Org-Committed to *y*, from *y* 's request a S-Commitment derives of *x* to *y* to do *a*. ;
- so, in fact, *y* has *indirectly* Committed *x* to do *a*.

This kind of **Indirect Commitments**, as well as the Command Power and the faculty to decide the actions of other people, are obviously a crucial and very well known feature of Organizations. This feature depends on the Org-Commitments of the members, and on their being bound by obligations.

Let us, at this point, just mention a criticism about the "Language/Action view" of Organization [Winograd 1987]. According to this view, agents seem to be completely free to negotiate and establish any sort of Commitment with any partner, without any constraint provided by dependence and power relations, norms and procedures, pre-established plans and cooperations. For us, not only behavior in Organizations is bound by "external" procedures, norms and rules, but even in a purely contractual perspective --based on agents' Commitments and their mental attitudes-- we saw that the agent is not free to establish his specific and extemporary Commitments. He is bound by his *role* (previous Org-Commitments) and by the consequent expectations and duties. In this sense, there is *no Organization without Obligations*. For this reason too, the Commitments an agent establishes in his cooperative work in an Organization, are not all equivalent, nor can be handled in the same way. For example, it should be distinguished whether a Commitment is merely "personal" (either by friendship or by social exchange) or it is "by role" [Castelfranchi, Conte, & Cesta 1992].

Conclusions

AI current interest in group work and Organization basically stems from the introduction of computers and their role in supporting and mediating cooperation. What is then required is *a theoretical understanding of human cooperative activity and of Organizations*. We think that AI is right in identifying the core of this problem in the notion of Commitment and in the mental representations of the agents. However, as we tried to show, the current notions are insufficient for accounting for the link between individual mind and collectivity. In particular, we introduced the intrinsically relational notion of *Social Commitment* (which is neither mutual nor collective) as an intermediary between the Internal Commitment and the Collective one. We analysed many relevant aspect of this social relation, with particular attention to its normative ingredients (obligations, expectations). We think that a formalization of these notions, as well as of the notions of Organizational Commitment and of Role, could help both in the theory of Organizations and Groups, and in the computational supporting of cooperative work. Our attempt has been just to clarify some useful concepts and to propose *a descriptive ontology* for this domain. Current views of Organization risk to be too "subjective" and too based on communication. They risk to neglect, on one side, the objective basis of social interaction (dependence and power relations), and on the other side, its normative components. The notions we proposed try to relate also to these levels of analysis.

Acknowledgements. Preliminary versions of this paper were presented at the *AAAI Workshop'93 on "AI and Theories of Groups and Organizations: Conceptual and Empirical Research"*, Washington, D. C., July 11-15 1993, and at the 4th JURAD Meeting, Madeira, March 28-29 1994. I would like to thank both the participants in those workshops for the interesting discussions and my

Friends of the PSCS Project, Amedeo Cesta, Rosaria Conte, Maria Miceli for their precious comments.

This work as been finished in the framework of the ESPRIT III Working Group "*ModelAge*", No.8319.

References

Bond, A. H. 1989. Commitments, Some DAI insigths from Symbolic Interactionist Sociology. In Proceedings of the 9^ International AAAI Workshop on Distributed Artificial Intelligence. 239-261. Menlo Park, Calif. : American Association for Artificial Intelligence, Inc.

Bouron, Th. 1992. Structures de communication et d'organisation pour la cooperation dans un univers multi-agents. These de Doctorat, Universite' Paris 6.

Castelfranchi, C. 1990. Social Power. A point missed in Multi-Agent, DAI and HCI. In Y. Demazeau, and J. P. Muller eds. *Decentralized A. I.* Amsterdam: Elsevier. 49-62.

Castelfranchi, C. 1992. No More Cooperation Please! In search of the Social Structure of Verbal Interaction. In Ortony, A., Slack, J., and Stock, O. eds. *Communication from an Artificial Intelligence Perspective*, Heidelberg, Germany: Springer 206-227.

Castelfranchi,C., Miceli,M., and Cesta,A. 1992. Dependence Relations among Autonomous Agents. In E. Werner, and Y. Demazeau. eds. *Decentralized A. I. 3*, Amsterdam: Elsevier. 49-62.

Castelfranchi,C., Conte,R., Cesta,A. 1992. The Organization as a Structure of Negotiated and Non Negotiated Binds. In G. C. van der Veer, M. J. Tauber, S. Bagnara, and M. Antalovits eds. *Human Computer Interaction: Tasks and Organization.* Proceedings of the Sixth European Conference on Cognitive Ergonomics (ECCE6), Balatonfured, Hungary, September 6-11.

Castelfranchi, C., Cesta, A. 1993. On Rational Reasons for Dropping Commitment in Joint Activities. In Preproceedings of the IJCAI Workshop on "Conflict management", IJCAI'93, Chambery, France.

Cohen, P. R. and Levesque, H. J. 1990. Intention is Choice with Commitmentitment. *Artificial Intelligence* 42:. 213-261.

Conte, R., Miceli, M., Castelfranchi,C. 1991. Limits and Levels of Cooperation: Disentangling various types of prosocial interaction. In Y. Demazeau,and J. P. Muller. eds. *Decentralized A. I. 2.* Amsterdam: Elsevier. 147-157)

Conte, R., Castelfranchi, C. 1993. Norms as mental objects. From normative beliefs to normative goals. In Preproceedings of the AAAI Spring Symposium on "Reasoning about Mental States: Formal Theories & Applications", AAAI,Stanford, CA, March 23-25.

Conte, R. and Castelfranchi, C. 1995. *Cognitive and Social Action.* London: UCL press.

Fikes, R. E. 1982. A commitment-based framework for describing informal cooperative work. *Cognitive Science*, 6: 331-347.

Gasser, L. 1991. Social conceptions of knowledge and action: DAI foundations and open systems semantics. *Artificial Intelligence* 47: 107-138.

Jennings. N. R. 1993. Commitments and conventions: The foundation of coordination in multi-agent systems. *The Knowledge Engineering Review* 3: 223-50.

Jones, A. J. and Porn, I. 1985. Ideality, sub-ideality and deontic logic. *Synthese* 65:295-318.

Levesque, J. H., Cohen, P. R., and Nunes,J. H. 1990. On Acting Together. In Proceedings of AAAI-90. Menlo Park, Calif.: American Association for Artificial Intelligence, Inc.

Rao,A. S., Georgeff, M. P., and Sonenberg, E. A. 1992. Social Plans: A preliminary Report. In E. Werner, and Y. Demazeau. eds. *Decentralized A. I. 3.* Amsterdam: Elsevier.

Searle, J. R. 1969. *Speech Acts.* Cambridge: Cambridge University Press.

Sichman,J. S., Conte,R., Castelfranchi,C., and Demazeau,Y.,A. 1994. Social Reasoning Mechanism Based On Dependence Networks. In Proceedings of ECAI'94. Amsterdam: ECCAI, August 8-12.

Singh, M. P. 1991. Social and Psychological Commitments in Multiagent Systems. In Preproceedings of "Knowledge and Action at Social & Organizational Levels", AAAI Fall Symposium Series, 104-106 Menlo Park, Calif.: American Association for Artificial Intelligence, Inc.

Werner, E. 1988. Social Intentions. In Proceedings of ECAI-88, Munich, WG, 719-723. ECCAI.

Winograd, T. A. 1987. Language/Action perspective on the Design of Cooperative Work. In *Human-Computer Interaction* 3, 1: 3-30.