Collective trust and normative agents

CLARA SMITH[1], *Facultad de Informática and IRI, University of La Plata, Argentina*
E-mail: csmith@info.unlp.edu.ar

ANTONINO ROTOLO, *CIRSFID and Law School, University of Bologna, Italy*
E-mail: antonino.rotolo@unibo.it

Abstract

In this paper we analyze the notion of collective trust within a multi-modal setting. We argue that collective trust is a scalable concept and therefore definable in qualitatively distinct levels or strengths. We show possible connections between different forms of group trust and the emergence of obligations within groups of agents. In particular, the notion of collective trust appears to be strong enough to entail forms of delegation that may have a deontic connotation: the trust a group puts in an agent's performance of an action yields to the result that it becomes obligatory in the interest of the group that the agent performs this action.

Keywords: collective trust, norms, agents.

1 Introduction

Besides the individual trust that some agent may have towards others or with respect to some situations, more complex patterns are often involved in social interaction. Suppose that three agents x, y and z agree that some goal A should be jointly achieved. Some kind of coordination among them is of course required, but, minimally, such a multi-lateral agreement at least implies that each agent trusts that the others jointly intend to achieve A, and also believe in that. This simple agreement thus presumes a relatively elaborated collective trust background.

This aspect is important in the law, where trust protection plays an important role [1], [2], [3]. The task of protecting trust is related to the problem of providing tools to support legally valid interactions between any kind of agents and/or to legally ground contractual transactions [1], [2]. Indeed, trust protection is strongly implemented especially when agents' beliefs seem reasonable or when trustees' behaviour induces trusters' reliance [1], [2], [3]. However, in multi-lateral agreement it is often the case that such reliance is mutual and this fact is relevant for trust protection. In particular, if agent x breaks group trust with regard to A, trust deception must be checked against the fact that x was supposed by the others to intend A, and x believed so.

The importance of trust has also been recognized as crucial in multi-agent systems (MAS, see e.g. [4]). In [5] a cognitive model of individual trust is given in terms of necessary mental ingredients which settle under what circumstances an agent x trusts another agent y with regard to an action or state-of-affairs, i.e. under which beliefs and goals an agent

[1]Partially funded by ANPCyT, Argentina. Project PICT 2006–00842.

doi:10.1093/jigpal/jzp076

delegates a task to another agent. Trust is then, in [5], understood as the mental counterpart of delegation of tasks, which can be weak or strong. Weak delegation means that there are delegation situations which do not suppose any agreement, deal or promise at all, nor which yield to rights. Stronger forms of delegation are the basis for promises, commitments and conventions. Weak and strong delegation support the idea of trust having different degrees of intensity which in their turn reflect the fact that, either alone or grouped, agents communicate, negotiate and coordinate themselves in a wide variety of ways, from altruism and cooperation to strategic situations with deep disagreement or strict business.

The challenge of formalizing the various kinds of group trust lies not only in the characterization of its different degrees. Collective forms of trust, when based upon strong delegation, turn out to be essential for modelling notions such as communal commitments and obligations, and special forms of lawful collective responsibilities, among others. For example, one central issue related to how groups work in the legal domain is the special structure of rules regarding the Roman notion of solidary obligation (e.g., Argentinean Civil Code arts. 699 to 717). An obligation is ''solidary'' (*in solidum*) when, by virtue of the title of the obligation, there is the need to satisfy the whole object of the obligation. Solidary obligations can be either conventional (usually contractual) or legal (e.g. parents are both *in solidum* responsible for the damages occurred due to their child's behaviour). The structure of a solidary obligation raises a ''common front'' either of debtors, creditors, or both. Each member of the front can behave, in principle, as creditor or debtor of the whole. The distinct individual links in a solidary obligation do not subsist alone nor are isolated; they are concentrated and converge in what gives the solidary obligation its unary structure: what happens to one actor is propagated, in principle, to the others [6]. So far, what an actor x does as a member of the front is imputed to the other members, as long as x acts in representation of a communal interest. Solidary obligations appear then to be a natural scenario for the emergence of common trust in the representative's behaviour.

The present paper focuses on one main MAS issue: how groups work, especially, how trust works within a group, and what its consequences are.

We investigate which are the correct formalizations for some essential aspects of complex kinds of group trust. One technique for doing this is shifting from single-agent notions to multi-agent ones through collective operators, an approach which we follow here. The approach is also followed by [7] where an agent's social attitudes are the starting point for an understanding of collective intentions in teams of agents. It is also the account followed in [8] where standard single-agent techniques are extended to multi-agent versions for proving complete axiomatizations of the modal systems **K**, **T**, **S4**, **KD45** and **S5**.

The paper is organized as follows. In Section 2 we establish a logical context in which to define group trust. We take as a basis the single-agent definition given in [5] which uses individual intentions, beliefs and goals. In Section 3 we shift to group trust and argue that it is scalable. We ground different levels of group trust not only on the kind of delegation of tasks involved (weak or strong) but also on the type of attitudes of agents: motivational or informational. This means that agents maintain an intentional attitude towards their environment, and that agents have information about the environment, respectively. We depict several scenarios for illustrating our intuitions and definitions; for each scenario we give a logical representation that corresponds to the intended level of trust involved. We make reference to Argentinean/Italian law when outlining institutions such as solidary obligations, irregular societies, a general provision to compensate damages, and norms in a code of conduct. We also take two examples from [3] regarding the withdrawal of pre-contractual

negotiations and the duty to disclose. In Section 4 we argue that a special type of group trust, which we call collective trust, implies such a strong form of delegation that it may have a deontic connotation: the trust a group puts in an agent's performance of an action A yields to the result that it becomes obligatory in the interest of the group that the agent performs A. Section 5 sketches a possible extension of the previous account of collective trust. Basically, the idea is to describe different degrees of collective trust within Fitting's [18] many-valued modal logic setting. Concluding remarks are provided in Section 6.

2 Language and Logical Background

The aim of this paper is to construct different forms of collective trust by

- taking as starting point [5]'s definition of individual trust, and
- working, with a few minor adjustments, within the logical framework developed by [7].

We need a language to talk about actions, and mutual and common beliefs and intentions. We briefly outline our standard language in this section.

On account of [9]'s advice, we use [7]'s multi-modal approach for dealing with agents' attitudes, but integrated by adding an agency operator (see also [2]). Hence, we work with a finite set of agents $G=\{x, y, z...\}$ and a countable set of atomic propositional sentences usually denoted by p, q, r … . Complex expressions are formed syntactically from these in the usual inductive way using $\bot$ and T (for *false* and *true*), standard Boolean connectives, and the unary modalities we describe next.

We use $\mathrm{Goal_x}$ A to mean that agent x has goal A, where A is a proposition. Propositions reflect particular state-of-affairs, as in [7]. $\mathrm{Int_x}$ A is meant to stand for ''agent x has the intention to make A true''. Intentions within the area of Cooperative Problem Solving (CPS) are viewed as inspiration for goal-directed activities and referred to as agents' motivational attitudes. The informational (known as doxastic or epistemic) modality $\mathrm{Bel_x}$ A represents that agent x has the belief that A. The $\mathrm{Does_x}$ A operator is to be understood in the same sense given in [10] to represent successful agency i.e. x indeed brings about A. To simplify technicalities, we will assume that in expressions like $\mathrm{Does_x}$ A no modal operators can occur in the scope of $\mathrm{Does_x}$: A will denote any behavioural actions concerning only conducts of agents, such as withdrawal, inform, purchase, payment, etc.

As classically established in [7], Goal is a $\mathbf{K}_n$ operator, while Int and Bel are, respectively, $\mathbf{KD}_n$ and $\mathbf{KD45}_n$ ones. The logic of Does, instead, amounts to the following schemata [17]:

$$\mathrm{Does_x A} \rightarrow \mathrm{A}$$
$$(\mathrm{Does_x A} \wedge \mathrm{Does_x B}) \rightarrow \mathrm{Does_x}(\mathrm{A} \wedge \mathrm{B})$$
$$\neg \mathrm{Does_x T}$$

Finally, we define below single-agent trust by reframing in the above multi-modal logic the account provided in [5]:

$$\mathrm{Trust}^{x}_{y}\mathrm{A} \leftrightarrow \mathrm{Goal_x(A)} \wedge \mathrm{Bel_x Does_y A} \wedge \mathrm{Int_x}(\mathrm{Does_y A} \wedge \neg \mathrm{Does_x A}) \wedge \mathrm{Bel_x Int_y A} \wedge \mathrm{Goal_x Int_y A} \quad (1)$$

which stands for: agent x trusts in agent y w.r.t A whenever x has goal A, and x believes that y indeed does A, and x has the intention of not doing A himself but of y doing it instead, and x has the goal that y intends A and x also believes in that.

We use (**1**) as a basis for shifting to collective definitions of trust (Section 3). For doing this we also use [7]'s mutual intention operator M-Int^{G}(A) which is meant to be true if every agent in group G intends A, everyone in G intends that everyone in G intends A, etc:

$$\text{M-Int}^{G}(A) \leftrightarrow \text{E-Int}^{G}(A \wedge \text{M-Int}^{G}(A)).$$

Here, the E-Int operator stands for ''every agent in G intends'' [7]. We also adopt the common belief operator C-Bel^{G} which is defined as follows:

$$\text{C-Bel}^{G}(A) \leftrightarrow \text{E-Bel}^{G}(A \wedge \text{C-Bel}^{G}(A))$$

The common belief that A is meant to be true if everyone in G believes in A, and everyone in G believes that everyone in G believes in A, etc. In this context, the E-Bel^{G} operator means ''every agent in G has the belief that'' [7].

[7]'s semantics for these and other iterative concepts is quite complex but well-established in the literature, and so we omit here the details. All the usual definitions of common belief and common knowledge have a graph-theoretical interpretation. We refer the reader to [7] and only recall here the valuation clauses for M-Int^{G} and C-Bel^{G}. Let a world v be G_I-reachable from a world w if there is a path of length ≥ 1 in [7]'s Kripke model M from w to v along accessibility arrows I_i which are associated with members i's in G. Then, if V is the valuation function in M,

$$V(w, \text{M-Int}^{G}(A)) = 1 \text{ iff } V(v, A) = 1 \text{ for all } v \text{ which are } G_I\text{-reachable from } w.$$

Analogously for C-Bel^{G}(A) using a similar notion of G_B-reachable.

It should only be noted that our analysis will not substantially add anything new to [7]'s semantic techniques: those standard constructions can be directly applied to all iterative trust concepts described in the remainder. However, the notion of agency we use here amounts to a non-normal modal system and so requires to reframe [7]'s original semantics into an equivalent one based on multi-relational models [17]. The advantage of this semantics is to preserve the basic intuition and structure of Kripke models, and allows us not to deviate from [7]'s account.

Definition 1. A multi-relational frame F is a structure

$$F = < G, W, \{B_i\}_{i \in G}, \{G_i\}_{i \in G}, \{I_i\}_{i \in G}, \{\mathcal{D}_i\}_{i \in G} >$$

where:

- G is the finite set of agents $\{x, y, z, \ldots\}$;
- W is a set of situations, or points, or *possible worlds*;
- $\{B_i\}_{i \in G}$ is a set of accessibility relations wrt Bel, which are transitive, euclidean and serial;
- $\{G_i\}_{i \in G}$ is a set of accessibility relations wrt Goal;
- $\{I_i\}_{i \in G}$ is a set of accessibility relations wrt Int, which are serial;
- $\{\mathcal{D}_i\}_{i \in G}$ is a family of sets of accessibility relations D_i wrt Does, which are pointwise closed under intersection, reflexive and serial [17].

Definition 2. A multi-relational model is a structure $M = <F, V>$ where:

- F is a multi-relational frame, and
- V is a valuation function defined as follows:
  1. standard Boolean conditions;
  2. $V(w, \text{Bel}_i\ A) = 1$ iff $\forall v$(if $w\ B_i v$ then $V(v, A) = 1$);
  3. $V(w, \text{Goal}_i\ A) = 1$ iff $\forall v$(if $w\ G_i v$ then $V(v, A) = 1$);
  4. $V(w, \text{Int}_i\ A) = 1$ iff $\forall v$(if $w\ I_i v$ then $V(v, A) = 1$).
  5. $V(w, \text{Does}_i\ A) = 1$ iff $\exists D_i \in \mathcal{D}_i$ such that $\forall v(w\ D_i v$ iff $V(v, A) = 1)$.

According to how we defined our multi-modal logic, these models—which trivially embed [7]'s into those described in [17]—do not require to reframe [7]'s semantic conditions for iterative concepts such as mutual intention and common belief: for the normal modalities we set a unique accessibility relation for each agent, while in the case of Does we need to define a set of relations for each of them. For the same reasons, soundness and completeness for the resulting logic follow directly from [7] and [17].

3 Delegation and Trust

In the following subsections we outline three qualitatively distinct, and progressively stronger, levels of group trust. We call them joint trust, reliance and collective trust. We establish definitions in the form of abbreviations of formulas within the language already provided, as we did with (**1**). (Other approaches consider trust to be a modality, as e.g. [12] where modalities capture the trust an agent has in the validity and completeness of a formula.)

The criteria for obtaining progressively stronger forms of group trust are based on whether suitable mutual intentions and common beliefs occur within the group. Indeed,

- mutual intentions are usually adopted in the literature (see [7]) to capture the idea of *group coordination* in realizing those states of affairs which are the content of such intentions;
- common beliefs are rather meant to express *group expectations*, namely, group awareness with respect to those states of affairs corresponding to the content of such beliefs (see, once again, [7]).

In the simplest case, we do not have any of the above conditions: joint trust is a straightforward generalization of (**1**), based on the concept of weak delegation. For example, agent x weakly delegates when, being at the bus stop, s/he relies on another agent y to raise up her/his hand predicting that y will do this [5].

Joint trust ''awareness'' is slightly stronger, as it is to be understood as the cognitive perception of the condition that everybody trusts.

Reliance is placed one step above joint trust and joint trust awareness. It is a specific joint mental attitude that reflects the fact that agents trust and they intend all others to trust. For example, when waving arms in a stadium everybody trusts that the others trust that everybody raises arms and therefore the wave is shaped. Neither joint trust nor reliance presuppose any agreement or deal among agents; moreover, they do not yield to commitments or rights at all.

Finally, collective trust is common belief on a common intention. It implies such a strong form of delegation that, although it does not yield to rights in the strict sense, its betrayal

may raise an obligation to compensate in the interest of the trusters. For example, consider the members of an irregular society (i.e., one that it has not been registered in accordance with the registration procedures and requirements). These members are solidary debtors and/or creditors—w.r.t. third parties—of the negotiations of the society[2]. Suppose that a member, due to a stipulation made in the name of the society, receives a full payment. Members trust that consequences arising from this payment are propagated to all of them, usually in conformity to a general principle of distribution.

The final subsection shows logical connections among joint trust, reliance and collective trust, and straightforward generalizations for each of the given definitions.

3.1. Joint Trust: Union of Individual Trust

Let G be a set of agents. It is simple to take a broader view of (**1**) to get everyone's trust as a sum of individual intentions and beliefs:

$$\text{Jtrust}^{\text{G}}_{\text{y}}\ \text{A} \leftrightarrow \bigwedge_{\text{i}\in\text{G}} \text{Trust}^{\text{i}}_{\text{y}}\ \text{A} \tag{2}$$

which is meant to stand for "G members jointly trust in agent y w.r.t. A".

Example 1. Suppose that agent y is at the bus stop, and people are standing not at the bus stop but near y, predicting that y will raise his hand and stop the bus. Let us call the people G. The fact can be modelled as $\text{Jtrust}^{\text{G}}_{\text{y}}$ (StopBus).

Discussion. Agents may be completely independent from the others, and therefore just look as a group from the outside. Nevertheless, it can be assumed that in this form of weak delegation trusters are comfortable in being dependent on the trustee(s) [5], [11]. (**2**) does not reflect any collective intention nor agreement at all. (Technically, it does not include any collective operator.) Jtrust captures a specific joint mental attitude which appears only on the grounds that members in G predicted that y will stop the bus. It is worth emphasizing that one may (unconditionally) trust. In Example 1 there is no commitment at all between y and G, nonetheless G trusts. Now let x be a member of G. If the bus approaches and y does not seem to be interested (e.g. he is waiting for another bus), and fearing x that he will miss the bus, x will probably raise his hand and stop himself the bus. Hence, in this case Jtrust holds until member(s) in G decide not to delegate to y the task of stopping the bus (temporalizing this scenario is beyond the scope of this paper).

For an akin and more lawful scenario regarding artificial agents, let us consider an on-line auction. The corresponding action to the one of raising the hand to stop the bus can be seen as making the opening bid(s) for a good.

3.2. Jtrust Awareness and Reliance

We next give definitions for two standpoints towards Jtrust. The first one reflects the fact that the group is aware of Jtrust, i.e. there is a cognitive perception of the condition that everybody trusts. The second one suggests a coordination situation encompassing intentions.

[2]Case-law of the Argentinean Supreme Court related to art. 699 of the Argentinean Civil Code and to art. 288 of the Commercial Code. In [6], pp. 513.

Being aware of Jtrust is:

$$\text{IJtrust}^{\text{G}}_{\text{y}}\ \text{A} \leftrightarrow (\text{Jtrust}^{\text{G}}_{\text{y}}\ \text{A}) \wedge \text{C-Bel}^{\text{G}}(\text{Jtrust}^{\text{G}}_{\text{y}}\ \text{A}). \tag{3}$$

Jtrust awareness (IJtrust) intends to capture a positive introspection towards Jtrust. The idea is to reflect that the group is aware of the group awareness of Jtrust. The philosophy here is rather pragmatic; (**3**) can be used to compute if everyone in G trusts in y and everyone in G believes that everyone in G trusts in y, and everyone in G believes that everyone in G believes that everyone in G trusts in y, etc. (**3**) may be useful for different applications or uses. For example, if G is aware of its Jtrust (i.e. it has IJtrust) members may find themselves in a position of moving to a trustworthy collective attitude. Suppose the following scenario:

Example 2. It is Mary's birthday. With a view to buying a coat Mary has already seen and described as nice, everyone in the office gives some money to y, a co-worker who is going downtown, delegating to y the search and purchase of the coat.

Discussion. Example 2 is a coordination situation. There is a collective attitude. The group relies on y on the basis of an instrumental reason (y going to downtown) rather than on y's intentional acting with goodwill, even when acting with goodwill may also be the case [5]. The instrumental reason suffices for entrusting y with their money. Moreover, not only co-workers trust in y, but also they intend all others in the group to have the associated intention to trust in y, and the intention that all members have the intention to trust in y, and so on. This reading meaning is rather strong, but we believe it fully capture the idea of group coordination. We call it Reliance and write it as:

$$\text{Rel}^{\text{G}}_{\text{y}}\ \text{A} \leftrightarrow (\text{Jtrust}^{\text{G}}_{\text{y}}\ \text{A}) \wedge \text{M-Int}^{\text{G}}(\text{Jtrust}^{\text{G}}_{\text{y}}\ \text{A}) \tag{4}$$

which is meant to stand for ''group G relies upon agent y w.r.t. A''. Everyone in G trusts in y (by (**2**)) and there is the mutual intention that everyone in G intends everyone in G to trust in y, etc. The key notion is that of mutual-intention, which captures a primary sign of a ''common front'' among members: they act in a coordinated way.

Note that in (**4**) we identify the minimal conditions for reliance to be present, leaving apart aspects that may also play a role in raising some degree of confidence, as commitments might do. Usually, commitments or promises induce a trustworthy atmosphere; nonetheless trust as a mental attitude may appear in situations where commitments are not clearly or explicitly settled, as it may be the case in both Example 1 and Example 2.

As we did for Example 1, an analogous situation among software agents for Example 2 is, e.g., an on-line purchase.

3.3. Collective Trust

Next follows the qualitatively stronger version of collective trust within our account. Collective trust is a complex collective informational and motivational attitude which involves more than a common intention (as Reliance does), it is a common belief on a common intention. Therefore, there is full confidence. This means strong delegation of tasks within the group. The definition is:

$$\text{Ctrust}^{\text{G}}_{\text{y}}\ \text{A} \leftrightarrow (\text{Rel}^{\text{G}}_{\text{y}}\ \text{A}) \wedge \text{C-Bel}^{\text{G}}(\text{Rel}^{\text{G}}_{\text{y}}\ \text{A}) \tag{5}$$

which settles that group G has common trust in y w.r.t. A whenever G relies upon y (i.e., there is a common intention) and there is the common belief that G relies upon y w.r.t. A (i.e. the group is aware of a common expectation towards a common intention). This in turn means that everyone in G believes that G relies on y (G is aware of its reliance on y), and everyone in G believes that everyone in G believes that G relies on y, etc. We illustrate with the following example.

Example 3. Every December in almost each neighbourhood of a university town, youth bands build up street-puppets filled with fireworks, which are to be burned on New Year's day. The task usually demands several days and a hard teamwork. The town administration has recently institutionalized a competition and settled an award for the best figure. Bands' custom establishes that figures ought to be watched and protected day and night. This because a very common practice (and offence) is to burn other band's figure before New Year's day by sending one's band member (i.e. a saboteur) to burn other's figure while its guardians are unaware (usually sleeping). The obvious consequence of successful sabotages is the exclusion of opponents from the competition.

Discussion. Betraying collective trust by e.g. negligence or insolvency surely has unwanted consequences, which are in principle spread to all members. This aspect is what makes collective trust strong: members have expectations, they believe in the success of the enterprise. Assume that s, band A's saboteur, is sent to burn band B's figure. We get the fact $\mathrm{Ctrust}_{s}^{\mathrm{A}}$ Burn. Band A believes on s's success; it is more interested in s's success rather than observing or controlling her/his action from the outside. If s fails, A's members are deceived in their beliefs, and it is probably the case that such a failure puts some or all members in a position (e.g. an internal/moral position) of not being themselves saboteurs in a next sabotage. Moreover, s may be punished for his negligence. Differently, in Example 2, a common interest has linked distinct individual interests to the *better* fulfilment of the former interest, but members do not necessarily believe that y will effectively buy the coat. If y fails to purchase the coat then their common intention is dissolved and somehow split; which means that members can reassume their positions to buy their own single presents.

3.4. Straightforward Generalizations for group Trust

It is easy to generalize (**3**), (**4**) and (**5**) to get:

$$\mathrm{Jtrust}_{\mathrm{H}}^{\mathrm{G}}\ \mathrm{A} \leftrightarrow \bigvee_{\mathrm{j}\in\mathrm{H}} (\mathrm{Jtrust}_{\mathrm{j}}^{\mathrm{G}}\ \mathrm{A}), \tag{6}$$

$$\mathrm{Rel}_{\mathrm{H}}^{\mathrm{G}}\ \mathrm{A} \leftrightarrow ((\mathrm{Jtrust}_{\mathrm{H}}^{\mathrm{G}}\ \mathrm{A}) \wedge \mathrm{M\text{-}Int}^{\mathrm{G}}(\mathrm{Jtrust}_{\mathrm{H}}^{\mathrm{G}}\ \mathrm{A})), \tag{7}$$

and

$$\mathrm{Ctrust}_{\mathrm{H}}^{\mathrm{G}}\ \mathrm{A} \leftrightarrow ((\mathrm{Rel}_{\mathrm{H}}^{\mathrm{G}}\ \mathrm{A}) \wedge \mathrm{C\text{-}Bel}^{\mathrm{G}}(\mathrm{Rel}_{\mathrm{H}}^{\mathrm{G}}\ \mathrm{A})). \tag{8}$$

The expressions above model joint trust, reliance and collective trust between and within finite groups of agents, respectively. It should be clear that awareness of joint trust can also be settled wrt groups when y is a group in (**2**). (**6**) reads intuitively as ''G trusts in H wrt A''. This is a natural generalization for joint trust: everybody in G weakly delegates A in any member of H. Next follow two examples for illustrating (**6**) and (**8**).

Example 4 (Bus stop example, revisited). Suppose that y is at the bus stop, now in a group we call H. Assume that people in the group we called G is near H. Then the following formula means that everyone in G trusts that someone in the group at the bus stop (H) and to which y belongs will raise her/his hand: $\text{Jtrust}^{\text{G}}_{\text{H}}$ (StopBus). Here, if anyone in H raises her/his hand then everyone in G profits from the bus being stopped (everyone in H does too).

Example 5. Group C, a group of creditors, and group D, a group of debtors, stipulated that D will reimburse in a fixed date an amount of money C lend to D.

Discussion. In solidary obligations the general rule is that the payment made by one of the co-debtors sets the others for free (e.g. arts. 699, 706 of the Argentinean Civil Code). This principle raises the following implication:

$$(\text{Does}_{\text{d}}\text{Reimburse}) \rightarrow (\text{Ctrust}^{\text{D}}_{\text{d}}\ \text{Released}),\ \text{d} \in \text{D}.$$

Moreover, creditors trust, in general, that they obtain their portion from the one who received the payment (e.g. according to art. 708 ACC.). This principle raises, e.g., the implication

$$(\text{Does}_{\text{d}}\text{Reimburse} \wedge \text{Does}_{\text{c}}\text{ReceivePayment}) \rightarrow \text{Ctrust}^{\text{C}}_{\text{c}}\ \text{Divide},\ \text{c} \in \text{C}, \text{d} \in \text{D}.$$

Betraying Ctrust may allow the members of D and C to raise actions and defences, as we address in Section 4.

Table 1 summarizes the concepts defined throughout Sections 2 and 3.

3.5. Logical Connections among Degrees of Trust

It is simple to find out that reliance implies joint trust and that collective trust implies reliance. Example 1 shows that joint trust does not imply reliance (i.e., it does not hold that $\text{Jtrust}^{\text{G}}_{\text{y}}\ \text{A} \rightarrow \text{Rel}^{\text{G}}_{\text{y}}\ \text{A}$). This because in (**2**) mutual intention is not required to hold. Similarly, it can be stated that (**4**) does not yield to beliefs and then reliance does not imply collective trust (it does not hold that $\text{Rel}^{\text{G}}_{\text{y}}\ \text{A} \rightarrow \text{Ctrust}^{\text{G}}_{\text{y}}\ \text{A}$).

TABLE 1. Formulas and their intended meanings.

Formula	Meaning
$\text{Trust}^{\text{x}}_{\text{y}}\ \text{A}$	Agent x trusts in agent y w.r.t. A
$\text{Jtrust}^{\text{G}}_{\text{y}}\ \text{A}$	G members jointly trust in agent y w.r.t. A
$\text{Jtrust}^{\text{G}}_{\text{H}}\ \text{A}$	G members jointly trust in group H w.r.t. A
$\text{IJtrust}^{\text{G}}_{\text{y}}\ \text{A}$	G members are aware of their joint trust in y w.r.t. A
$\text{Rel}^{\text{G}}_{\text{y}}\ \text{A}$	G relies upon agent y w.r.t. A
$\text{Rel}^{\text{G}}_{\text{H}}\ \text{A}$	G relies upon group H w.r.t. A
$\text{Ctrust}^{\text{G}}_{\text{y}}\ \text{A}$	G has common trust in y w.r.t. A
$\text{Ctrust}^{\text{G}}_{\text{H}}\ \text{A}$	G has common trust in group H w.r.t. A

These statements help to highlight from a logical point of view that trust can be conceived in different qualitative levels. This leads us to considering that levels of trust among interactive groups or among groups and individuals may vary while they interrelate.

Example 6. ([3] Case 3, p. 236): Firm F, an organization of accountants, enters negotiations with the owner of an office block, Group G, with the view to renting space for their new office. In the course of these negotiations, F asks for various works to be done. Before an appointment fixed by both parties for signing the contract, Firm F finds an offer in equally suitable premises but at a lesser rent, and refuses to go further with G.

Discussion. Consider that, when negotiations begin, trust levels for F and G might be at a first stage, say Jtrust. Assume that when F asks for various work to be done G's trust raised let's say to a reliance level. Finally, G's trust probably increased to a Ctrust level when the appointment for signing the contract was fixed. Group G is induced by Firm F to believe in the future conclusion of a contract. Although contracts are ruled by the general principle of autonomy, the law can be seen as frequently recognising or supporting some kind of trust situations as in this case, where the very fact that one party trusts is the reason why the law provides for its protection through the generation of an obligation. In such situations, it is said that trust comes first and law comes after [1]. (For the anglo-saxon position regarding pre-contractual negotiations see, e.g., [19, 20])

The fact that G reached a qualitatively high trust setting (Ctrust) raises the issue of the emergence or not of an obligation for firm F to compensate the withdrawal of pre-contractual negotiations. That is, Ctrust in this example seems to imply such a strong form of trust that it yields to the actor's obligation of performing what it is entrusted to. If Firm F, the counterparty, abandons the negotiations without justification, it violates standards of correctness and it is therefore bound to restore the damage suffered by the other party[3], group G.

Example 6 raises another issue: whether Ctrust has its own range of values. And if so, which value trust should reach to immediately trigger a compensation.

In the following sections we deal with a deontic connotation for Ctrust with a view to capture a schema powerful enough for the immediate triggering of compensations for trust deception (Section 4). Further, we address an extension to the two-valued approach which provides a basis for a multi-valued concept for trust (Section 5). We do not give necessary or sufficient conditions for triggering compensations, but rather establish the underlying formal structure of a multi-agent normative system that may handle strategies related to compensations.

4 A Deontic Connotation for Collective Trust

Assume that we have an action and deontic logic which allows us to formally specify normative systems. That is, in addition to the operators for beliefs, goals, intentions, and agency we also have deontic operators such as O and O^G for representing obligations. O is the deontic operator for generic obligations, meaning "it is obligatory that" [2] [9], and O^G is a relativised obligation operator which is meant to stand for "it is obligatory in the interest of G that" (see e.g. [13]).[4] In what follows, we restrict ourselves to the use of the relativised

[3] Nonetheless, we cannot assume that the law provides a general duty to restore damage to trusters [11, 14].

[4] We assume that $O^G\ A \leftrightarrow \bigwedge_{i \in G} O^i\ A$ holds.

deontic operator, which we assume is characterised by the usual $\mathbf{KD}_n$ semantics. Notice that therefore, correspondingly, the multi-relational frame in Definition 1 is to be extended with the set of accessibility relations $\{O^i\}_{i\in G}$ w.r.t. the relativised obligations, which are serial. Also, let us assume that we deal with a theory of trust which enables us to write propositions of the form (**1**) to (**8**) using the language in Section 2.

A previous development of the idea of an operator with a deontic connotation (or, more generally speaking, a concept with such connotation) appears in [14]. There, the notion of commitment has a deontic connotation by defining the commitment to perform α as a meta-action with the result that it becomes obligatory to perform α. The property for which the definitions are devised is $\mathrm{INT}(\alpha)\rightarrow[\mathrm{COMMIT}(\alpha)]\mathrm{O}(\alpha)$, meaning that if one intends to perform the action denoted by α then one is obliged to perform that action after having committed one's self to do so. For providing a lawful support to Ctrust, we devise the following schema:

$$(\mathrm{Ctrust}_y^G\ A)\rightarrow O^G\ (\mathrm{Does}_y A) \tag{9}$$

which aims to reflect the lawful force of Ctrust, relativized to groups.

Following [1], (**9**) can be understood as a standard of behaviour (good faith behaviour, or faithful behaviour). This standard can be identified with reference to social or group norms, to correctness, or reasonableness. The duty to compensate damages is the only protection which can be provided to the truster in ''behaviour-trust'' situations [1] e.g. situations where the one who puts reasonable trust on somebody is lawfully protected in her/his interests. Thus, by virtue of (**9**) and provided that the normative system we are considering settles a provision to compensate damages, e.g.,[5]

$$O^G\mathrm{Does}_y A\wedge(\neg\mathrm{Does}_y A)\rightarrow O^G(\mathrm{Does}_y\mathrm{Compensate}) \tag{10}$$

we can e.g. deduce $O^G\ (\mathrm{Does}_y\mathrm{Compensate})$.

Example 7. ([3] Case 16, p. 447): *d* is a doctor who treats *p* for a severe back injury. He prescribes a medication which considerably slows down *p*'s reactions. *d* forgets to tell *p* about these effects of the drug. After *p* has taken the drug, *p* causes an accident while driving his car which is attributable to his poor reaction. Can *p* institute a claim against *d*?

Discussion. In Italy, the Code of Conduct of the medical profession specifies the duties of doctors in relation to the prescription of medicines. These rules do not have the force of law, being contained in neither primary nor secondary legislation. But they are relevant to the question whether or not doctors act diligently. Assume that one of the rules in the Code of Conduct is: Prescribe → Inform, meaning that the prescription of medicines involves the task of informing its effects. Let P be a group of potential patients and D a group of doctors; let $p\in$ P, $d\in$ D. Given the fact that patients trust in doctors' overall conduct, we have by (**9**):

$$\mathrm{Ctrust}_D^P(\mathrm{Prescribe}\rightarrow\mathrm{Inform})\rightarrow O^P\mathrm{Does}_d(\mathrm{Prescribe}\rightarrow\mathrm{Inform}).$$

Now, prescribing the medicine but failing to inform its effects implies failing to do what it is obligatory in the interest of P and the violation of (**9**). Hence by (**10**) we deduce $O^P\ (\mathrm{Does}_d$ Compensate).

[5]For example, on the basis of art. 2043 of the Italian CC, or art. 1068 of the Argentinean CC.

Other interesting examples somehow amount to the reverse case of what we have already discussed.

Example 8. Let us slightly reframe Example 6. Suppose that agent i, member of Firm F is fully entrusted by F to enter into negotiations with the owner j of a Group G, with the view of renting the above mentioned space for their new office. Before signing the contract, i, on behalf of Firm F, finds an equivalent offer but at a lesser rent. Group G is strongly induced by i to believe in the future conclusion of the contract, and for this reason G declines to sign a contract with a third party. On account of this fact, if Firm F abandons negotiations without justification it violates standards of pre-contractual correctness and it can be under the obligation to restore the damage suffered by G. Here, collective trust in F play another role: that of providing grounds to ascribe a collective liability towards F. So we could not only have

$$((\mathrm{CTrust}^{\mathrm{G}}_{\mathrm{F}}\mathrm{A}) \wedge (\neg \mathrm{Does}_{\mathrm{y}}\mathrm{A})) \rightarrow \mathrm{O}^{\mathrm{G}}(\mathrm{Does}_{\mathrm{y}}\mathrm{Compensate}),$$

but also

$$((\mathrm{CTrust}^{\mathrm{G}}_{\mathrm{F}}\mathrm{A}) \wedge (\neg \mathrm{Does}_{\mathrm{y}}\mathrm{A})) \rightarrow \mathrm{O}^{\mathrm{G}}\ (\bigvee_{\mathrm{k}\in\mathrm{F}} \mathrm{Does}_{\mathrm{k}}\mathrm{Compensate}).$$

Example 8 depicts a scenario where a peculiar type of ''solidary'' obligation is arising from the group trust. Indeed, what an agent does as a member of the front is imputed to the other members, as long as such an agent acts in representation of a communal interest. In this example, however, this mechanism is used to ascribe collective liability.

5 Further Extensions: Degree of Collective Trust in Many-valued Modal Logics

Examples 5, 6, 7, and 8 raise the question of which values for Ctrust may trigger a compensation (or reparation, or reimbursement). Instead of expecting Ctrust to amount to be true in order to engender an obligation to compensate (which should be understood as the standard two-valued case) strategies w.r.t. other amounts of Ctrust can be defined for raising such compensation. In other words, triggering a compensation/reparation may require that collective trust goes beyond a certain threshold of intensity.

In the remainder, we briefly outline a framework in which a many valued (m-v) definition for Ctrust can be devised. Therefore, (**9**) and (**10**), for example, get applicable to cases only when this threshold is crossed, and this can be anticipated when we aim at designing a team or system behaviour.

The amount of Ctrust needed clearly depends on circumstances and varies, for instance, from the class of emergency cases (for example, a vessel in the sea or other situations where life is involved) and other complex cases where agents, roles or communication are missing and therefore low values for Ctrust should trigger obligations, to cases and situations where team experience implies a good or effective teamwork and therefore high amounts of Ctrust are needed to be betrayed to trigger a compensation. Here we confine ourselves to only show how to extend the (traditional multi-modal multi-agent) two-valued approach of Sections 2 through 5 with a m-v approach.

5.1. Many-valued Trust, Version I

The following m-v account is a direct generalization of standard modal logic [18, Version I]. We outline here the basics of [18]'s account.

We work on a space of truth values that constitutes a finite lattice; this means that conjunction and disjunction are understood as natural notions, but negation may not. In the simplest case, implications are allowed and interpreted by the partial ordering of the lattice but in general cannot be nested. We will assume that the lattice is distributive, which allows for nestable implications [18].

This m-v account retains the notion of possible world semantics while allowing formulas to have values in a many-valued space τ. Initially, we have $\tau = <L, \wedge, \vee, 0, 1>$ an arbitrary finite distributive lattice, that constitutes a space of truth values. L is a non-empty set of elements (the 'values') and the lattice ordering is written $\leqslant$. There are two binary operations $\vee$ and $\wedge$ read 'join' and 'meet', and two nullary operations, 0 and 1 which are the bottom and top elements of τ but we will call them respectively *false* and *true*. It holds that *false* $\neq$ *true*. In m-v logics it is not always possible to define within the language every truth value (such as $p \vee \neg p$ defines *true* and $p \wedge \neg p$ defines *false*), so we explicitly add to our language in Section 2 propositional constants for each member of τ. Hence these ''names'' can be used as atomic formulas.

We proceed as follows. First, we provide an intuitive m-v reading of the single-agent operators of our language, namely Bel, Goal, Int, and Does.[6] Then we outline the semantic structures needed to support the m-v theory for trust.

- The Bel_i case. There are natural reasons why to valuate beliefs in a m-v space. We do not always accept things as true, nor false. For example, there are situations where we 'hardly' believe or 'slightly' believe others. We may accept a state-of-affairs, we agree up to a certain point with others' ideas. We do not ourselves guarantee every debt or default whenever we are asked to, probably because we do not always rely on the principal debtor's behaviour. We are sometimes not fully aware of how things are going on or carried out. Bel_i A will therefore give us the amount of belief agent i puts in A. Different amounts of beliefs may be relevant to model agents' intuitions, non-rational beliefs, or beliefs based on agents' previous experience.
- The Goal_i case. Goals can be ranked along different dimensions, for example, relevance. Thus Goal_i A gives us the relevance A has for agent i. It is important to notice that highly-ranked goals do not always commit us to act. To illustrate this, think of the goals we cannot directly achieve because we must be 21 years old, or we must first get a degree, or earn enough money to accomplish them. Differently, we sometimes act in accordance with low-ranked goals because they are simple actions, or because we are altruist people, etc. (strategies are not handled within agents in our model).
- The extension for the Int_i case is straightforward: intentions are, in our framework, a subset of goals.
- The Does_i case. Recall we restrict ourselves to single behavioural actions (Section 2). The intuition behind a m-v account for Does is to capture the degree of success of the performance of an act A. Consider this performance as a unit (i.e., a unit of action in a unit of temporality, which is in our account tacitly presupposed). Then a faultless performance of A coincides with *true*, while lower valuations mean A has

[6] For brevity, we omit here to discuss the case of directed obligations.

been carried out although not perfectly accomplished (e.g., A has been 'almost', or 'roughly', or 'more or less' accomplished). For instance, the qualification we obtain in an examination is a degree of performance. Let us illustrate with another example from international public law. Suppose A is the act of signing a convention among states. One state party may agree with some provisions but disagree with others. So, it may consent to be bound by some provisions but formulate reservations to others, or object any of them, or maintain a pending entry into force. Therefore, the provisional application of such convention by the state may not be considered a full performance. As other analogous examples, consider partial information giving, and partial payment.

Let us see how to adjust the semantics of Section 2 for the purposes of this m-v multimodal logic. First of all, it should be noted that we are not required to change the notion of frame provided in Definition 1 [18]. Hence, every multi-relational frame is also a m-v multi-relational frame. What we need to change is how to evaluate formulas.

Definition 3. *A m-v multi-relational model is a structure* $M=<F,V>$ *where:*

- F is a multi-relational frame, and
- V is a valuation function that maps atomic formulas and worlds to τ.[7] Let w,v $\in$ W; then the mapping V is extended to all formulas in the usual way with the following special conditions, which stand for meet operations on τ:

1. $V(w, Bel_i\ A) = \wedge\{V(v, A), wB_iv\}$,
2. $V(w, Goal_i\ A) = \wedge\{V(v, A), wG_iv\}$,
3. $V(w, Int_i\ A) = \wedge\{V(v, A), wI_iv\}$,
4. For the Does case let us keep in mind that, for every agent *i*, *D* returns a set of binary relations, which gives us:

$$V(w, Does_i A) = \bigwedge_{D_j \in D_i} \{\wedge\{V(v, A), wD_jv\}\}$$

where the D_j are the different binary relations in *D*.

It is worth mentioning the status of the success condition $Does_i\ A \rightarrow A$ in this m-v account. This schema is at the core of the logic of agency; if i brings about A then A is the case. Now, by considering the structure of $\tau=<H, \wedge, \vee, \rightarrow, 0, 1>$ we have a Heyting algebra in which '$\rightarrow$' is interpreted as a m-v implication that semantically fits Does success condition. This because $V(Does_i\ A \rightarrow A)$ is the largest element c of H (in the lattice sense, i.e., it is a relative pseudo-complement [21]) such that $V(Does_i\ A \wedge c) \leq V(A)$, which suits the intuition that one can not bring about A better than A itself.

Provided the extension, it is plain to see that we directly get a many-valued definition for Trust (**1**), and also for Jtrust (**2**). This because (**1**) and (**2**) in this m-v version become a meet of values in τ. Although clear from the algebraic point of view, the intuition behind the intended meaning needs some comments. To illustrate, let us focus in (**1**): according to (**1**), the values assigned to each modal operator in Definition 3 are combined to define a value for $Trust^x_y$. This leads to combine values assigned to goals, beliefs, intentions, and actions, which have quite different readings. The resulting m-v value for $Trust^x_y$ is intuitively meant to stand for: ''the relevance goal A has for agent *x*, along with the amount up to which *x*

[7] Note that in a two-valued version of τ, V collapses to the usual valuation function.

believes in certain degree of success of agent y's performance of A, along with the relevance it has for x his intention regarding a certain degree of success of y's performance of A rather than its own performance of A, along with the relevance it has for x the goal that y intends A to a certain degree, along with the amount up to which x believes in that intention of y''. (Again, see the structure of the formula in (**1**).)

Following, E-Int^{x}_{y} A has a straightforward m-v version, based on the same grounds:

$$V(w, \text{E-Int}_G A) = \wedge_{i \in G}\{V(v, \text{Int}_i A)\} = \wedge_{i \in G}(\wedge\{V(v, A), wI_i v\}).$$

The extensions of M-Int and C-Bel to m-v versions are also direct, as the former one is an iterative definition built on top of several applications of E-Int and the latter is an iterative definition built on top of Bel_i. For example, for all $w, v \in W$,

$$V(w, \text{C-Bel}^G(A)) = \wedge\{V(v, A) \text{ such that } v \text{ is } G_B\text{-reachable from } w\}.$$

All this work gives many valued definitions for reliance and collective trust in (**3**) and (**4**) respectively.

5.2. Many-valued Trust, Version II

We next reframe [18]'s example within the bus stop scenario we gave in Section 2. This seems appropriate with a view to motivate the usefulness of this alternative m-v account before getting to the technical details regarding such a complex collective attitude.

Example 9 (Bus Example, revisited). Agents a and b are near the bus stop. Suppose both are asked to pass a judgement on the truth of the statement of stopping the bus, in various situations. A natural truth value to work with is a four-valued one, meaning: both say yes, a says yes but b says no, b says yes but a says no, and both say no. Then these truth values can be identified with subsets of $\{a, b\}$. Now, two kinds of opinions are possible: that a statement A is true in situation w, and that situation w should be taken into account. The first type of opinion reflects the assignment of a truth value to A at w. The second one amounts to a many-valued accessibility relation. Suppose that there are three situations, w, w_1 and w_2. Let w be ''this'' world, e.g. ''today'', let w_1 be ''tomorrow under a rainstorm'', and let w_2 be ''tomorrow carrying the children back home''. Let A be ''stopping the bus''. Suppose agents a and b say w_1 should be considered, but only a says that w_2 should be. Assume also that only b says A would be true in situation w_1, and nobody says A would be true in situation w_2. How should ''the bus will be stopped, no matter what'' that is, $\Box$A is to be evaluated in this world?

Discussion. Fitting [18] says that it should be clear that the value of $\Box$A is what is common to all alternative situations. As far as w_1 goes, everyone says it should be considered but only b says A is true there. Thus intuitively from agent a we get a no and from b a yes, so w_1 contributes with $\{b\}$. For w_2, b does not say it should be considered at all. Then w_2 cannot serve as a counterexample for b so w_2 counts as a yes for b. For agent a w_2 should

be considered but A is false there, so w_2 counts as a no for a. Thus w_2 also contributes $\{b\}$ and so $\Box$A is given the value $\{b\}$ in this world.

This semantics captures the one given by Version I and in addition it allows models to have many-valued accessibility relations. Intuitions behind this Version II account of each single-agent operator in our trust theory follow:

- The Bel_i case. We do not always accept things as true, nor false; we may accept or be aware of them in a wide variety of degrees, provided the context. Suppose three situations, w, w_1 and w_2. While being in world w agent i may consider that s/he should believe that A in situation w_1 but say s/he would hardly believe that A while being in situation w_2.
- The $Goal_i$ case. Goals can be ranked along the dimension which refers to how favourable they are w.r.t. different situations. For example, agent i may easily find A as a goal in situation w_1 but s/he would barely consider A as a goal in situation w_2. Nonetheless, favourable situations for goal achievement do not necessarily commit to act. Such a void reaction may be due to other factors which are independent of the situation (e.g., moral/legal prohibitions).
- The extension for Int_i A is straightforward: recall that intentions are, in our framework, a subset of goals.
- The $Does_i$ case. The intuition is to capture the worthiness of the performance of an act A w.r.t. a given situation. In the convention signing example, while being in this world one state party may consider to fully agree with the terms in the convention in situation w_1 (e.g. a peace and welfare situation) although it may consent to be bound by some provisions but formulate reservations to others in situation w_2 (e.g., a war or invasion situation).

Now, the semantics. In essence, differences regarding version I appear in the definition of the valuation function. We provide below only some guidelines to adjust the models introduced in Section 2.

Basically, accessibility relations for beliefs, goals, and intentions are mappings with signature $W \times W \to \tau$, conceived of as many valued relations between possible worlds. In the case of agency, for every agent i in G, D_i is a function $W \times Pow(Pow(W)) \to \tau$ corresponding to the many valued accessibility relations wrt agency.

Once again, V is a valuation function that maps atomic formulas and worlds into τ under to the condition that members of τ map into themselves. In particular, let w,v $\in$ W. The mapping V is then extended to all formulas in the usual way, with the following special conditions, which stand for meet operations on τ:

$$V(w, Bel_i A) = \wedge\{wB_i v \to V(v, A)/v \in W\}.$$
$$V(w, Goal_i A) = \wedge\{wG_i v \to V(v, A)/v \in W\}.$$
$$V(w, Int_i A) = \wedge\{wI_i v \to V(v, A)/v \in W\}.$$

Regarding Does, we get:

$$V(w, Does_i A) = \wedge\{wD_i n_j \to V(n_j, A)/n_j \in Pow(W)\}, \text{ where:}$$
$$V(n_j, A) = \underset{k \leq j}{\wedge} V(v_k, A), v_k \in W.$$

We moved the single-agent operators to a multi-valued space. It is not hard to see here, too, that we directly get a many-valued definition for Trust (**1**) and for Jtrust (**2**).

6 Concluding Remarks

In this paper we investigated some forms of collective trust and suggested a perspective on how these forms of trust can be logically related to the emergence of obligations within groups of agents. Our starting point has been the definition of individual trust proposed in [5]. We argued that the basic ingredients of trust can be captured within a modal approach. This approach, which is also widely accepted when collective attitudes are considered, proved useful in identifying trust settings in multi-agent systems, each corresponding to a different degree of group confidence. Finally, we discussed, with special attention to the legal domain, the relation between collective trust (which is a form of trust based on mutual belief and strong delegation of tasks) and the emergence of normative conventions within groups of agents.

Our contribution provides evidence that minimal adjustments are required for existing frameworks such as those in [7] to deal with a theory of common trust and norms. We think that this makes our proposal valuable, since we show that frameworks such as [7]'s can be easily extended to cover highly structured scenarios involving trust.

Clearly, other intuitions regarding group trust can be devised, for example one based on agents combining their beliefs. Assume that if agent x trusts in agent z w.r.t. A then agent y also trusts in z w.r.t. A provided that x and y belong to the same group and under the assumption that A can be derived from the set of communal goals. Therefore, group trust in this context intuitively means ''if someone in the group trusts, then everyone trusts'': *from* Trust_z^x A *and* $\{\mathrm{Goal}\ g_i\} \vdash$ A *then infer* (Trust_z^x A $\rightarrow$ Trust_z^y A), with x,y $\in$ G. This is an intuitive inference, grounded on a pool of trusts. It exploits the view used to justify the intuition regarding distributed knowledge [8]. We can use it as a basis for building collective forms of trust as we did throughout Section 3.

In addition, patterns such as Ctrust can be used to capture some aspects of group conventions such as those based on promises. For doing this we need some previous definitions. First, under appropriate conditions, promises can be defined as: Prom_y^x A $\leftrightarrow$ ($\mathrm{Does}_x\mathrm{Bel}_y\mathrm{Int}_x$ A), meaning that agent x promises A to y when agent x makes y believe that x has the intention to A. From the success property of Does it is easy to check that promises generate expectations: ($\mathrm{Does}_x\mathrm{Bel}_y\mathrm{Int}_x$ A) $\rightarrow$ ($\mathrm{Bel}_y\mathrm{Int}_x$ A). Promises raise commitments when they introduce x's acknowledgement of y's ''entitlement'' [5]: when x promises A to y, the usual feedback is that at least y believes that x will fulfil what promised, and x is aware of this. Then we can define commitments as: Comm_y^x A $\leftrightarrow$ (Prom_y^x A $\rightarrow$ $\mathrm{Bel}_x\mathrm{Bel}_y\mathrm{Int}_x$ A). Finally, within our group trust theory we can set up the principle that groups have the collective expectation that commitments are fulfilled within the group: Ctrust_x^G ((Comm_y^x A) $\rightarrow$ A), x,y $\in$ G.

To summarize, as it is well-known, the modal approach is open to critiques. In particular, while feasible modal systems for collective attitudes (based on normal logics such as **KD45**) pose the omniscience problem, other weaker systems minimize it but are of less utility due to their weaknesses. In [2] we have proposed a logical methodology for modelling individual trust and good faith that avoids some of these difficulties. How to extend this system to cover collective attitudes and trust is a matter of future research.

Finally, in Section 5 we outlined how to extend the definition of group attitudes (and so collective trust) within [18]'s many-valued setting. It seems that this approach can be a promising method for introducing degrees of collective trust in [7]'s setting. However, a number of technical issues are still open, one of them being how to adjust [7]'s completeness proof to cover the many-valued case, a task which is far from obvious. Also this point is left to future research.

References

[1] D. Memmo, G. Sartor, G. Q. di Cardano. Trust, Reliance, Good Faith and the Law. *Trust Management*, 2003.

[2] A. Rotolo, G. Sartor, and C. Smith. Good Faith in Contract Negotiation and Performance. *International Journal of Business Process Integration and Management*, forthcoming, 2009.

[3] S. Whittaker, R. Zimmermann. *Good Faith in European Contract Law*. Cambridge University Press, 2000.

[4] C. Castelfranchi and Y.-H. Tan (eds.). *Trust and Deception in Virtual Societies*. Kluwer, 2001.

[5] C. Castelfranchi and R. Falcone. Social Trust. A Cognitive Approach. In [4].

[6] J.J. Llambías. *Civil Code With Annotations. Vol. II-A, Obligations in general*. Abeledo-Perrot, Bs. As. 1979.

[7] B. Dunin-Keplicz, R. Verbrugge. Collective Intentions. *Fundamenta Informaticae* XX, 2002, pages 1–25.

[8] J. Halpern, Y. Moses. A Guide to Completeness and Complexity for Modal Logics of Knowledge and Belief. *Artificial Intelligence 54*, 1992, pages 311–379.

[9] A.J.I. Jones, M. Sergot. A Logical Framework. In *Open Agent Societies: Normative Specifications in Multi-Agent Systems*, Wiley, 2007.

[10] D. Elgesem. The Modal Logic of Agency. *Nordic Journal of Philosophical Logic*, 1997.

[11] R. Tuomela. A Collective's Rational Trust in a Collective's Action. Understanding the Social II: Philosophy of Sociality, *Protosociology* 18–19:87–126. 2003.

[12] R. Demolombe, C.-J. Liau. A Logic of Graded Trust and Belief Fusion. 4th *Workshop on Deception, Fraud and Trust in Agent Societies*, Montreal, 2001.

[13] H. Herrestad, C. Krogh. Deontic Logic relativised to Bearers and Counterparties. In: J. Bing and O. Torvund eds., *Anniversary Antology in Computers and Law*. TANO 1995.

[14] F. Dignum, J.-J.Ch. Meyer, R. Wieringa, R. Kuiper. A Modal Approach to Intentions, Commitments and Obligations: Intention plus Commitment yields Obligation. In: M. Brown and J. Carmo eds., *Deontic Logic, Agency and Normative Systems*. Springer Verlag, Berlin, pp. 80–97, 1996.

[15] A. Artosi, G. Governatori, A. Rotolo. On the Logical Nature of Counts-as Conditionals. *LEA 2004*. Gedit, 2004.

[16] A.J.I. Jones, S. Firozabadi. On the Characterization of a Trusting Agent. Aspect on a Formal Approach. In [4].

[17] G. Governatori, A. Rotolo. On the Axiomatisation of Elgesem's Logic of Agency and Ability. *Journal of Philosophical Logic*, 2005, 34 (4), pp. 403–431.

[18] M C. Fitting. Many-valued Modal Logics. *Fundamenta Informaticae* 15, 1991, pages 235–254.

[19] H. MacQueen. Good Faith in the Scots law of contract: An undisclosed principle? *Good Faith in Contract and Property Law*. Hart, 1999.
[20] S. Whittaker, R. Zimmermann. Good Faith in European Contract Law: Surveying the Legal Landscape. In [3].
[21] S. Burris, H. P. Sankapannavar. *A Course in Universal Algebra*. Springer, 1981.