

Graded Trust

Robert Demolombe

Institut de Recherche en Informatique de Toulouse
France

* robert.demolombe@orange.fr

Abstract. After a brief analysis of several trust definitions a common pattern is exhibited which takes the form of a truster’s belief about the regularity of some trustee’s property. That leads to a definition of graded trust in terms of two independent components: graded belief and graded regularities, where grades take qualitative levels. This idea is formalized in the framework of classical modal logics. After an informal discussion of the axiom schemas and inference rules of the selected logic, a formal definition of its proof theory and of its model theory are given. Finally, the main features of this approach are compared with other proposals for the formalization of qualitative graded beliefs, in particular the Spohn’s approach.

1 Introduction

There are many definitions of trust, nevertheless most of them agree on the fact that trust is essentially a mental attitude of an agent, the truster, with regard to another agent, the trustee. This attitude involves truster’s beliefs, and also, in some definitions, other features like truster’s goal.

This concept of trust has been formalized in a quantitative framework, like probabilities [12], or in a qualitative framework, like formal logic [3, 8, 10, 6]. In formal logic most of the proposals deal with non graded frameworks. That is, either the truster trusts, or does not trust, the trustee.

In this paper a new framework for qualitative graded trust is proposed in modal logic. Two guidelines have been followed in designing this logic: the first one is to be compatible with most of the definitions of the concept of trust, though there is no consensus in this area, and the second one is to be compatible, as far as possible with a quantitative formalization, in particular with probability theory.

In the following sections we start with a brief survey of some of the most well known definitions of trust and a common core is exhibited. In section 3 it is shown that the notion of graded trust involves two components. This requires two independent grades that are called “graded beliefs” and “graded regularities”. The following section 4 is devoted to the analysis and definition of an appropriate modal logic; the first subsection is about the proof theory and the second one is about the model theory. In the last section our proposal is compared to related works, and some conclusions are presented.

* This work has been partially supported by the French Agence Nationale de la Recherche under contract ANR-06-SETI-006.

2 About trust

In [10] (see also [11]) A.J.I. Jones surveys several definitions of trust. He points out that for some authors, like M. Bacharach, and D. Gambetta [1], trust is a truster's expectation of an action to be performed by the trustee. More specifically for T. Rea [16] this expectation is about the trustee's competence and his fulfilment of all fiduciary obligations. In [9] K. Giffin, add to trust definition the fact that the truster's expectation is related to some truster's objective. Then, after a deep analysis of concrete scenarios Jones concludes that the minimal constituents of the core of trust definition should be defined in terms of truster's beliefs about some regularity and conformity properties satisfied by the trustee, and that truster's goal may also be present in the truster's attitude, but that this is not necessarily the case.

C. Castelfranchi and R. Falcone in [3] offer a different analysis, based on cognitive science, where they argue that the truster's goal is a constitutive part of trust definition, and they integrate other features in their definition, in particular truster's dependence about the trustee.

In [6] (see also [5]) R. Demolombe adopting a simpler trust definition has presented a classification of the different kinds of properties the truster may ascribe to the trustee. This classification is briefly recalled below in order to show that they all share a common formal pattern which can also be found in most of the other definitions, even if these definitions cannot be reduced to these patterns. The classification is defined in terms of epistemic, dynamic and deontic properties. Some examples are presented below.

Sincerity. Agent i trusts agents j about his sincerity about p iff i believes that IF j informs i about p ($Inf_{j,i}p$), THEN j believes p (Bel_jp).

In formal terms: $Bel_i(Inf_{j,i}p \Rightarrow Bel_jp)$.

Competence. Agent i trusts agents j about his competence about p iff i believes that IF j believes p (Bel_jp), THEN p holds.

In formal terms: $Bel_i(Bel_jp \Rightarrow p)$.

Vigilance. Agent i trusts agents j about his vigilance about p iff i believes that IF p holds, THEN j believes p (Bel_jp).

In formal terms: $Bel_i(p \Rightarrow Bel_jp)$.

Cooperativity. Agent i trusts agents j about his cooperativity about p iff i believes that IF j believes p (Bel_jp), THEN j informs i about p ($Inf_{j,i}p$).

In formal terms: $Bel_i(Bel_jp \Rightarrow Inf_{j,i}p)$.

Ability. Agent i trusts agents j about his ability to bring it about that p iff i believes that IF j has attempted to bring it about that p (H_jp), THEN p holds.

In formal terms: $Bel_i(H_jp \Rightarrow p)$.

Obedience. Agent i trusts agents j about his obedience about the obligation to bring it about that p iff i believes that IF it is obligatory that j brings it about that p ($ObgE_jp$), THEN j brings it about that p (E_jp).

In formal terms: $Bel_i(ObgE_jp \Rightarrow E_jp)$.

Honesty. Agent i trusts agents j about his honesty with respect to the permission to bring it about that p iff i believes that IF j brings it about that p (E_jp), THEN it is permitted that j brings it about that p ($PermE_jp$).

In formal terms: $Bel_i(E_jp \Rightarrow PermE_jp)$.

In these definitions, formulas of the form: $\phi_j \Rightarrow \psi_j$ can be read: ϕ_j entails ψ_j .

More recently, E. Lorini and R. Demolombe [15] have formalized in modal logic trust definitions which are very close to those proposed in [3]. For instance, they have defined trust in positive action as follows: *i trusts j to do α with regard to his goal that ϕ if and only if i wants ϕ to be true and i believes that:*

1. *j, by doing α , will ensure that ϕ , and*
2. *j has the capacity to do α , and*
3. *j intends to do α*

It can be easily shown that conditions 1 and 2 have a conditional form. For instance, the condition 1 can be rephrased as: "i believes that if j performs the action α , then ϕ holds".

At the end of this analysis, our conclusion is that in almost all the trust definitions we find patterns of the form:

$$Bel_i(\phi_j \Rightarrow \psi_j)$$

where $\phi_j \Rightarrow \psi_j$ represents some j 's property that i ascribes to j .

3 Graded trust

In most of realistic situations it is an over simplification to say that a truster i either trusts, or does not trust, a trustee j . Rather, in informal terms, we say, for instance, that ***i has a limited trust in j***, or ***i's trust in j is high***. Then, we are faced to this **question:** "what is the meaning of such sentences?".

A **first answer** to the question, when trust is represented by a formula of the form $Bel_i(\phi_j \Rightarrow \psi_j)$, is that i is **uncertain** to be in a world where the set of ϕ_j worlds (the set of worlds where ϕ_j is true) is included into the set of ψ_j worlds (the set of worlds where ψ_j is true). For example, i may be uncertain about the fact j is sincere about p , that is, about the fact that in every circumstances where j informs i about p it is the case that j believes p .

Here, graded trust can be defined by the strength level of i 's belief about j 's sincerity. Notice that this "uncertainty" level refer to the validity of i 's beliefs, not to the completeness of i 's beliefs. In more formal terms graded trust can be represented by the formula: $Bel_i^g(\phi_j \Rightarrow \psi_j)$, and this formula can be read: *the strength level of i's belief about the fact that $\phi_j \Rightarrow \psi_j$ is true is g*, where Bel_i^g is used to denote a "graded belief".

A **second answer** may be that i believes that the set of ϕ_j worlds is "**partially included**" into the set of ψ_j worlds. In that case the fact that i 's trust in j 's sincerity is high can be interpreted as the fact that i believes that in almost all circumstances if j informs i about p , then j believes p .

The "inclusion level" of the set of ϕ_j worlds into the set of ψ_j worlds is called the "**regularity level**" of j 's attitude. This level is formally represented by the formula: $\phi_j \Rightarrow^h \psi_j$, and we also say that it represents a graded regularity. In that case graded trusts are represented by formulas of the form: $Bel_i(\phi \Rightarrow^h \psi)$.

Our proposal in this paper is that graded trust refers to **both** answers and that they should be represented by formulas of the form:

$$Bel_i^g(\phi \Rightarrow^h \psi)$$

whose intended meaning is that the strength level of i 's belief about the fact that ϕ entails ψ with a regularity level h is g .

4 A modal logic for graded beliefs

We have defined a formal logic for reasoning about graded trust in order to be able to derive the consequences of a set of assumptions that are supposed to represent a particular situation in a given application. This part of the logic is defined by its proof theory. It is complemented by its model theory whose objective is to formalize the meaning of the concepts, and their correspondent modalities. Roughly speaking, the model theory defines the meaning of the fact that a formula of this logic is true in a particular situation which is represented by a formal model.

4.1 Proof theory

The proof theory is presented progressively in order to explain the meaning and the justification of the inference rules and axiom schemas that have been chosen.

First, it is assumed that levels are represented by a finite non empty set of qualitative grades G , and that there is a total order on G represented by the relation: \leq . The highest level is denoted by *max*, and the lowest level is denoted by *min*.

To represent beliefs we have two modalities. The first one represents beliefs to which an agent has not assigned a strength level, for example, because he has not enough information to fix the grade of this belief. They are called "standard beliefs". The second one represents graded beliefs. Their notations and intuitive meanings are:

$Bel_i(\phi)$: i believes that ϕ is true.

$Bel_i^g(\phi)$: the strength level of i 's belief about the fact that ϕ is true is (exactly) g .

The logical connective \Rightarrow^h is a conditional in Chellas's sense [4]. As mentioned before it is used to represent graded regularities, and its intuitive meaning is:

$\phi \Rightarrow^h \psi$: ϕ entails ψ at the level h .

Semi formal analysis

For the modality Bel_i we have adopted as usual a KD logic (see [4]).

For the modality Bel_i^g we have the following inference rules and axiom schemas.

(U0) In $Bel_i^g(\phi)$, ϕ can be substituted by any logically equivalent formulas.

(U1) If ψ is a logical consequence of ϕ (i.e. $\vdash \phi \rightarrow \psi$), then if i has ascribed a strength level to his belief about ϕ and to his belief about ψ , then the level of ψ cannot be lower than the level of ϕ .

Notice that this rule does not impose that if i has ascribed a level to ϕ , he has necessarily also ascribed a level to ψ . The reason why we have this cautious rule is that it may be that ψ may contain sub formulas which are not relevant to ϕ . For example, it may be that the meaning of ϕ is that j is sincere, and the meaning of ψ is that j is

sincere or k is honest. In that example ψ is a logical consequence of ϕ . However, if i ignores who is k , i may have no opinion about the fact k is honest. Then, i cannot ascribe a level to ψ , although he knows that if he had to fix a level for ψ , it should be greater or equal to the level of ϕ .

It is also worth noting that we do not have the axiom schema: $Bel_i^g(\phi) \rightarrow Bel_i^g(\phi \vee \psi)$, because in most cases, if the level of $\phi \vee \psi$ is fixed, it is greater than g . This observation shows that Bel_i^g is not a normal modality, it is a classical modality.

(U2) If the levels of beliefs of the formulas ϕ_1 and ϕ_2 are fixed, then the level of their disjunction is the maximum of these two levels.

If the levels in graded beliefs would be interpreted in terms of probabilities, we would have that the probability of the disjunction is greater or equal to the maximum of the two probabilities. Then, the choice of (U2) is not compatible with probabilities, but it is as close as possible to probabilities when we are dealing with qualitative levels.

(U3) If the levels of beliefs of the formulas ϕ_1 and ϕ_2 are fixed, then the level of their conjunction is the minimum of these two levels.

The justification of (U3) is similar as the justification of (U2).

(U4) The strength level of i 's belief is unique for a given sentence.

(U5) Graded beliefs are consistent with standard beliefs. Notice that $Bel_i^{g_1}\phi$ and $Bel_i^{g_2}\neg\phi$ are consistent in a similar way as $g_1 = Pr(\phi)$ and $g_2 = Pr(\neg\phi)$ are consistent, provided $g_1 + g_2 = 1$.

(U6) If ϕ represents the formula which is believed at the minimum level and ψ is a standard belief, then ϕ implies ψ .¹

The intuitive idea is that there is no proposition that is believed at a lower level than the proposition that characterizes **all** i 's standard beliefs. In some sense this proposition is the most specific one which is believed (in a standard sense) by i .

(U7) If ϕ represents the formula which is believed at the maximum level and ψ is a standard belief, then ψ implies ϕ .

The intuitive idea is that there is no proposition that is believed at a greater level than the level of tautologies.

(U8) If ϕ is believed at the level g , then i believes that ϕ is believed at the level g .

This positive introspection axiom schema means that no level is ascribed by i to his evaluation of the level of a belief. If such a level would be ascribed, then one could ask the question: *what is i 's evaluation of this "second" order level?*, and we would be leaded to an infinite number of introspection levels, which is far to be intuitive.

Notice also that there is no justification to ascribe the maximum level to introspection beliefs because the maximum level is restricted to tautologies, while graded beliefs are contingent propositions.

(U9) If ϕ is not believed at the level g , then i believes that ϕ is not believed at the level g .

The justification of (U9) is similar as the justification of (U8).

For the conditional connective \Rightarrow^h we have the following inference rules and axiom schemas.

(R0) In $\phi \Rightarrow^h \psi$, ϕ and ψ can be substituted by logically equivalent formulas.

¹ We would like to thanks the anonymous referee who pointed out an error in the preliminary version of axioms schemas (U6),(U7) and (R2).

(R1) If ϕ entails ψ at the level h , then if ϕ holds, ψ holds "at the level" h .

This axiom schema can be easily understood if we think to its possible interpretation in terms of conditional probabilities. If $\phi \Rightarrow^h \psi$ is interpreted as: $h = Pr(\psi|\phi)$, if we have: $1 = Pr(\phi)$, we can infer that: $h = Pr(\psi)$, which can be rephrased as: $h = Pr(\psi|\top)$, and this formula can be seen as the interpretation of $\top \Rightarrow^h \psi$. That is why, in the following, ψ^h is used as a notation for: $\top \Rightarrow^h \psi$.

If we have $\phi_1 \Rightarrow^{h_1} \psi$ and $\phi_2 \Rightarrow^{h_2} \psi$ and $h_1 \neq h_2$, from ϕ_1 and ϕ_2 we can infer ψ^{h_1} and ψ^{h_2} , which contradicts the further unicity schema (R3). We have the same kind of contradiction with conditional probabilities if we have $h_1 = Pr(\psi|\phi_1)$ and $h_2 = Pr(\psi|\phi_2)$, and ϕ_1 and ϕ_2 are both true; because we get $h_1 = Pr(\psi)$ and $h_2 = Pr(\psi)$.

(R2) There exists a function F such that if $n = F(h_1, k_1, h_2, k_2)$, then, if ϕ entails ψ at the level h_1 , ψ entails θ at the level k_1 , ϕ entails $\neg\psi$ at the level h_2 and $\neg\psi$ entails θ at the level k_2 , then ϕ entails θ at the level n .

The reason why we have this axiom schema is that, in general, from: $((\phi \Rightarrow^{h_1} \psi) \wedge (\psi \Rightarrow^{k_1} \theta))$ we cannot infer what is the value of n such that: $(\phi \Rightarrow^n \theta)$, because there may be ϕ worlds that are θ worlds, and which are not ψ worlds. Notice that axiom schema (R2) is perfectly compatible with conditional probabilities if we accept some uniform distribution assumptions. In that case the form of F is: $n = (h_1 \times k_1) + (h_2 \times k_2)$.

(R3) The regularity level of ϕ entails ψ is unique.

Finally, it is worth noting that we do not have an axiom schema that allows us to infer from: $(\phi \Rightarrow^{h_1} \psi) \wedge (\phi \Rightarrow^{h_2} \theta)$, the value of h_3 such that: $(\phi \Rightarrow^{h_3} \psi \wedge \theta)$. The reason is that, for a given level of h_1 and h_2 , the set of ψ worlds and the set of θ worlds may be either disjoint or one may be included into the other one.

Formal definition

The proof theory is formally defined as follows.

The syntax of the language is defined as usual for a multimodal propositional logic (see [4]).

In addition to the inference rules and axiom schemas of classical propositional logic we have the following inference rules and axiom schemas.

Notations.

$$Forall(g, cond)F(g) \stackrel{\text{def}}{=} \bigwedge_{g \in G, cond(g)} F(g)$$

$$Exist(g, cond)F(g) \stackrel{\text{def}}{=} \bigvee_{g \in G, cond(g)} F(g)$$

$$\psi^h \stackrel{\text{def}}{=} \top \Rightarrow^h \psi$$

$Bel_i(\phi)$ obeys a KD system.

$$(U0) \text{ If } \vdash \phi \leftrightarrow \psi \text{ then } \vdash Bel_i^g(\phi) \leftrightarrow Bel_i^g(\psi)$$

$$(U1) \text{ If } \vdash \phi \rightarrow \psi \text{ then } \vdash Bel_i^g(\phi) \rightarrow \neg Exist(g', g' < g) Bel_i^{g'} \psi$$

$$(U2) \text{ If } g_3 = Max\{g_1, g_2\} \text{ then } \vdash Bel_i^{g_1}(\phi_1) \wedge Bel_i^{g_2}(\phi_2) \rightarrow Bel_i^{g_3}(\phi_1 \vee \phi_2)$$

$$(U3) \text{ If } g_3 = Min\{g_1, g_2\} \text{ then } \vdash Bel_i^{g_1}(\phi_1) \wedge Bel_i^{g_2}(\phi_2) \rightarrow Bel_i^{g_3}(\phi_1 \wedge \phi_2)$$

$$(U4) \vdash Forall(g_1, g_2, g_1 \neq g_2) \neg (Bel_i^{g_1}(\phi) \wedge Bel_i^{g_2}(\phi))$$

$$(U5) \vdash Bel_i^g \phi \rightarrow \neg Bel_i \neg \phi$$

$$(U6) \vdash (Bel_i^{min} \phi \wedge Bel_i \psi) \rightarrow (\phi \rightarrow \psi)$$

$$(U7) \vdash (Bel_i^{max} \phi \wedge Bel_i \psi) \rightarrow (\psi \rightarrow \phi)$$

- (U8) $\vdash Bel_i^g(\phi) \rightarrow Bel_i Bel_i^g(\phi)$
 (U9) $\vdash \neg Bel_i^g(\phi) \rightarrow Bel_i \neg Bel_i^g(\phi)$
 (R0) If $\vdash \phi \leftrightarrow \phi'$ and $\vdash \psi \leftrightarrow \psi'$ then $\vdash (\phi \Rightarrow^h \psi) \rightarrow (\phi' \Rightarrow^h \psi')$
 (R1) $\vdash (\phi \Rightarrow^h \psi) \rightarrow (\phi \rightarrow \psi^h)$
 (R2) There exists a function F such that if $n = F(h_1, k_1, h_2, k_2)$, then
 $\vdash ((\phi \Rightarrow^{h_1} \psi) \wedge (\psi \Rightarrow^{k_1} \theta) \wedge (\phi \Rightarrow^{h_2} \neg\psi) \wedge (\neg\psi \Rightarrow^{k_2} \theta)) \rightarrow (\phi \Rightarrow^n \theta)$
 (R3) $\vdash \text{Forall}(h_1, h_2, h_1 \neq h_2) \neg((\phi \Rightarrow^{h_1} \psi) \wedge (\phi \Rightarrow^{h_2} \psi))$

4.2 Model theory

The model theory gives a formal semantics to the concepts of standard beliefs, graded beliefs and graded regularities. Models are a particular sort of minimal conditional model as defined by Chellas (see [4], section 10.1). They are defined as a tuple M such that:

$$M = \langle W, \{B_i\}, \{B_i^g\}, \{R^h\}, v \rangle$$

In M , W is a set of possible worlds, $\{B_i\}$ is a set of functions: $B_i : W \rightarrow 2^W$, which assign to each world a set of worlds, $\{B_i^g\}$ is a set of functions: $B_i^g : W \rightarrow 2^{2^W}$, which assign to each world a set of sets of worlds, $\{R^h\}$ is a set of functions: $R^g : W, 2^W \rightarrow 2^{2^W}$, which assign to each pair formed with a world and a set of worlds, a set of sets of worlds, and v is a function: $v : ATOM \rightarrow 2^W$, which assigns to each atomic formula a set of worlds.

In this kind of models a proposition and the set of worlds where this proposition is true are identified.

The intuitive meaning of these functions can be seen through formal examples.

If $B_i(w) = X$, X is the set of worlds consistent with **ALL** the propositions believed by i in w . This set of worlds is usually characterized by an accessibility relation in models of normal modal logics.

If $B_i^g(w) = \{X_1, X_2\}$, the set of propositions believed by i in w at the level g is represented by the set of sets of worlds: X_1 and X_2 . That means that the set of propositions believed at the level g is represented by these two sets.

If $R^h(w, X) = \{X_2, X_4\}$, in w the set of propositions entailed at the level h by the proposition represented by X is represented by the set of sets of worlds X_2 and X_4 . These two sets can also be interpreted as two propositions.

Satisfiability conditions

For reasons that are explained in the comments about the satisfiability conditions for graded regularities, the truth value of a formula is defined in the context of a set of worlds X .

$M, X, w \models \phi$ can be read: ϕ is true in the world w , in the context X and in the model M .

This notion of truth relativized to a context is related to the standard notion of truth by the following condition.

$$M, w \models \phi \text{ iff } M, W, w \models \phi.$$

Let us assume that X is a subset of W .

$$\text{Notation: } |\phi|_X \stackrel{\text{def}}{=} \{w_1 : w_1 \in X \text{ and } M, X, w_1 \models \phi\}.$$

$M, X, w \models atom$ iff $w \in v(atom)$ and $atom$ is an atomic formula.
 $M, X, w \models \neg\phi$ iff $M, X, w \not\models \phi$.
 $M, X, w \models \phi \vee \psi$ iff $M, X, w \models \phi$ or $M, X, w \models \psi$.
 $M, X, w \models Bel_i\phi$ iff $\exists Y(Y = B_i(w) \text{ and } \forall w'(w' \in Y \Rightarrow M, Y, w' \models \phi))$.
 $M, X, w \models Bel_i^g\phi$ iff $\exists Y(Y = B_i(w) \text{ and } |\phi|_Y \in B_i^g(w))$.
 $M, X, w \models \phi \Rightarrow^h \psi$ iff $|\psi|_X \in R^h(w, |\phi|_X)$.

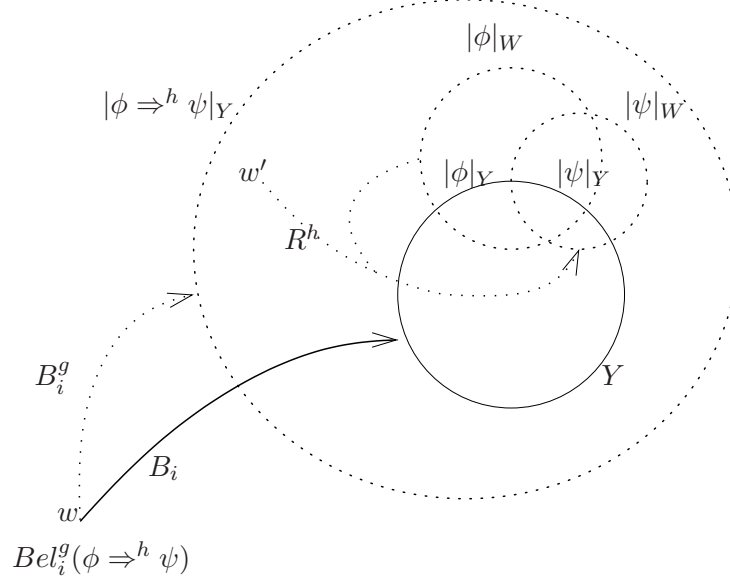


Fig. 1. Evaluation of a graded belief about a graded regularities.

Example. $M, X, w \models Bel_i^g(\phi \Rightarrow^h \psi)$ iff $\exists Y(Y = B_i(w) \text{ and } |\phi \Rightarrow^h \psi|_Y \in B_i^g(w))$. From $|\phi|_X$ definition we have: $|\phi \Rightarrow^h \psi|_Y = \{w' : w' \in Y \text{ and } M, Y, w' \models \phi \Rightarrow^h \psi\}$. From the satisfiability conditions we have: $M, Y, w' \models \phi \Rightarrow^h \psi$ iff $|\psi|_Y \in R^h(w', |\phi|_Y)$.

This example shows why formulas are evaluated with respect to a given context. Indeed, to evaluate to what extent i believes that the set of ϕ worlds is included into the set of ψ worlds, we have to restrict ϕ extension and ψ extension to the set of worlds which are consistent with all i 's beliefs in w , i.e. to the set of worlds Y , which is $B_i(w)$ (see figure 1). That is the reason why formulas are evaluated with respect to a given context. If a formula is not in the scope of some agent's beliefs, then the context is not restricted, i.e. the context is W .

Notice that, since graded beliefs are standard beliefs, the set of worlds $|\phi \Rightarrow^h \psi|_Y$ contains $B_i(w)$.

5 Related works

In [17] W. Spohn has defined a framework to represent graded beliefs in order to give a more satisfying account of rational epistemic changes. His final goal is to raise deterministic epistemology to a level as satisfying as probabilistic conditionalization in the field of non deterministic changes.

The framework, like in this paper, is defined by a set of possible worlds, and propositions are also identified to sets of worlds. A set of beliefs is represented by a set of worlds, called the "net content", the set of worlds which is included into all the believed propositions (in our framework this set of worlds is $B_i(w)$). The first idea is to represent belief changes with simple conditional functions (SCF) which collect all the possible changes of the net content of epistemic states brought about by all possible information. Then, a SCF g is a function from 2^W to 2^W . Spohn shows that the information represented by the SCFs can be represented by a well ordered partition (WOP) of W , where a WOP is a partition such that ordinals $0, \dots, n$ are assigned to each member of the partition. These ordinals are intended to represent the strength of **disbelief** of propositions represented by each partition. These members are denoted by E_0, \dots, E_n . The partition E_0 is the least disbelieved proposition and it represents the net content of an epistemic state.

According to these definitions a WOP represents a SCF iff for all non empty set of worlds A we have $g(A) = E_\beta \cap A$, where E_β is the least disbelieved partition that intersects A . On the basis of this correspondence between WOPs and SCFs it is shown that no SCF can appropriately represent epistemic changes, in the sense that it is possible to get the same epistemic change after getting and removing an information A , and that getting information A and then B leads to the same epistemic state as getting B and then A .

That is the reason why Spohn introduces the ordinal conditional functions (OCF). An OCF k is defined on a complete field of propositions and assigns to each non empty proposition an ordinal such that 0 is not obtained from an empty set, and the ordinal $k(w)$ is the same for all the worlds w in the same atomic proposition. Then, for any non empty proposition A , the ordinal $k(A)$ characterizes the least disbelieved world in A , i.e. $k(A) = \min\{k(w) : w \in A\}$. Notice that $k(A) = 0$ means that A is not believed to be false, and we may have both $k(A) = 0$ and $k(\neg A) = 0$.

Finally, to have the properties of reversibility and commutativity of epistemic changes a complementary parameter α is added to complement the values of k . The grade α characterizes the strength of $\neg A$. It is used to increment the value of $k(\neg A)$ after getting the information A . We have no room here to explain in detail how this parameter is defined and how it is used.

There are some commonalities with graded beliefs that have been presented here in the sense that qualitative grades are assigned to beliefs. For example, $Bel_i^{\min} A$ and $k(A) = 0$ have similar meanings. We also have $k(A \cup B) = \min\{k(A), k(B)\}$ which is very close to our axiom schema (U2). However, there are significant differences. The first one is that in our framework there is no need to assign a grade to all the propositions. The second one is that the meaning of the grades are different: in our framework they represent the strength of beliefs, while for Spohn they represent the strength of disbeliefs. Is there a one to one correspondence between each of them? The

answer is far to be obvious. The third one is that to define the OCFs we have to assign grades to all the worlds. We think that in a non trivial application domain where a world is defined by tens of atomic propositions, it is quite difficult to consider each world and to evaluate the appropriate grade for this world. May be a trick could be to "cluster" sets of worlds, and to assign to them the same grade, but that is exactly what we do if these sets are seen as propositions. The difference being that we do not request a "complete" assignment. Finally, in Spohn proposal there is no proof theory.

Several authors have taken inspiration in Spohn's proposal in the perspective of modeling belief changes (see, for example, C. Boutilier [2]). In [14] N. Laverty and J. Lang have explicitly integrated these ideas in a modal logical framework. They define a modality $B^i\phi$ whose intuitive meaning is that "the agent believes ϕ with strength i "² and whose satisfiability condition for a given OCF k is: $k \models B^i\phi$ iff $i \leq k(\neg\phi)$.

N. Laverty in [13] has defined a normal modal logic for graded beliefs where modalities $B^i\phi$ obey a $KD45$ system called $KD45_G$. The positive introspection (negative introspection is similar) axiom schema takes the form: $B^j\phi \rightarrow B^iB^j\phi$ which, in our view is questionable, as mentioned in section 4.1. The author shows how these modalities can be "translated" into the OCF framework. For a given OCF k we have: $k, s \models B^i\phi$ iff $\forall s'(k(s') < i \Rightarrow k, s' \models \phi)$.

In [15] E. Lorini and R. Demolombe have defined a normal modal logic for graded beliefs where $Bel^{\geq x}\phi$ can be read "agent i believes ϕ at least with strength x ". These modalities are interpreted by binary relations P_i^x , and $P_i^x(w)$ denotes the set of worlds accessible from the world w . These relations are structured by the constraint: if $y < x$ then $P_i^y(w) \subseteq P_i^x(w)$, which can be seen as a structure of spheres. From these modalities are defined modalities $Bel^x\phi$ whose meaning is that agent i believes ϕ at strength x . The satisfiability conditions for these modalities can be expressed as: $M, w \models Bel^x\phi$ iff $P_i^x(w) \subseteq |\phi|_W$ and $P_i^{suc(x)}(w) \not\subseteq |\phi|_W$. The correspondence with Spohn's OCF is defined by a translation of the set of spheres into an EOP. The significant point of this works is that these graded beliefs are integrated into a logical framework that defines different kinds of trust.

R. Demolombe and C. J. Liau in [7] have defined graded beliefs and graded trust in order to propose a solution to belief revision. They define modalities $B_i^\alpha\phi$ whose meaning is that agent i believes ϕ at the level α . These modalities are normal modalities. They also define classical modalities of the form $TV_{i,j}^\alpha\phi$ and $TC_{i,j}^\alpha\phi$, whose meaning are that agent i trusts agent j at the level α for being a valid (respectively complete) information source for ϕ . Here, a valid information source is an information source who is both sincere and competent, and a complete information source is defined in a dual way. The meaning of these graded trust definitions can be well understood with the axiom schemas: $TV_{i,j}^\alpha\phi \rightarrow (K_i Inf_{j,i}\phi \rightarrow B_i^\alpha\phi)$, and $TC_{i,j}^\alpha\phi \rightarrow (K_i \neg Inf_{j,i}\phi \rightarrow B_i^\alpha\neg\phi)$.

We have seen that most of the works dealing with graded beliefs have been done to formalize belief change. The few ones which consider graded beliefs for modeling graded trust identify the level of beliefs and the level of trust. The most significant difference with what we have proposed is the fact that we have considered that two

² In fact the meaning of this modality is that the agent believes ϕ with strength at least equal to i .

independent grades are involved in trust definition. The following example is intended to show why we need graded trust and graded regularities.

Let us assume that agent i has a low belief strength g about the fact that j is very competent with regard to p , because on the basis of 5 observations where j has believed p , in 4 situations it was the case that p was true. The strength of i 's belief is low because the number of observations is quite limited. If after a greater number of observations, say 5 additional observations, it is confirmed that j is very competent, the i 's belief strength g will be greater while the grade h of j 's competence remain the same. However, if for these 5 new observations it happens that j 's belief was wrong in 3 cases, i will believe that j has a moderate competence, that is that h is lower. Even if the grades g and h are not necessarily assigned on the basis of observations, from this simple example we can understand why we need two different grades to evaluate i 's trust about j 's competence.

6 Conclusion

A logical framework has been defined to represent graded trust in terms of two independent components: graded beliefs and graded regularities. We do not pretend that trust can be reduced to a set of components of this kind, but we have shown that they are included in most of the trust definitions.

A class of non normal modalities formalize graded beliefs, while standard beliefs obey a normal modal system. Graded regularities are also represented by non normal modalities. The model theory for these operators is defined in the framework of minimal conditional models "a la Chellas". It is possible not to ascribe a grade to all the standard beliefs. For example, in the financial domain, it may be that i believes that the trader j is competent, but i does not have enough background in the domain to assign a grade to j 's competence.

We have accepted a framework which has limited logical properties but offers more flexibility for further specializations. For example, it is not imposed to assign a grade to every believed proposition, as it is the case in Spohn's framework. If for some specific reason one would like to impose such constraint, it would not be a great difficulty to add correspondent constraint in the logic.

In the future we want to analyze to what extent the framework could be adapted to a quantitative analysis in terms of probabilities. A possible direction to be investigated would be to interpret graded beliefs as subjective probabilities, and graded regularities as objective conditional probabilities. For example, sentences of the form: $Bel_i^g(\phi \Rightarrow^h \psi)$, could be interpreted as: $Bel_i Pr_{sub}(Pr_{obj}(\psi|\phi) = h) = g$.

Another direction for future works is to go deeper in the analysis of mathematical properties of the logical framework. In particular the constraints to be imposed to the models in order to validate the axiom schemas must be analyzed carefully. For example, (U2) is valid if we impose the constraint: (CU2) If $X_1 \in B_i^{g_1}(w)$, $X_2 \in B_i^{g_2}(w)$ and $g_3 = Max\{g_1, g_2\}$, then $X_1 \cup X_2 \in B_i^{g_3}(w)$, and (U5) is valid if we impose the constraint: (CU5) If $X \in B_i^g(w)$, then $B_i(w) \cap X \neq \emptyset$.

Finally, we believe that the constraint to have a total order on the set of grades could be easily relaxed if that is required in a particular domain.

References

1. M. Bacharach and D. Gambetta. Trust as type detection. In C. Castelfranchi and Y-H. Tan, editors, *Trust and Deception in Virtual Societies*. Kluwer Academic Publisher, 2001.
2. C. Boutilier. *Conditional Logics for Default Reasoning and Belief Revision*. PhD thesis, University of Toronto, 1992.
3. C. Castelfranchi and R. Falcone. Social trust: a cognitive approach. In C. Castelfranchi and Y-H. Tan, editors, *Trust and Deception in Virtual Societies*. Kluwer Academic Publisher, 2001.
4. B. F. Chellas. *Modal Logic: An introduction*. Cambridge University Press, 1988.
5. R. Demolombe. To trust information sources: a proposal for a modal logical framework. In C. Castelfranchi and Y-H. Tan, editors, *Trust and Deception in Virtual Societies*. Kluwer Academic Publisher, 2001.
6. R. Demolombe. Reasoning about trust: a formal logical framework. In C. Jensen, S. Poslad, and T. Dimitrakos, editors, *Trust management: Second International Conference iTrust (LNCS 2995)*. Springer Verlag, 2004.
7. R. Demolombe and C-J. Liau. A logic of graded trust and belief fusion. In C. Castelfranchi and R. Falcone, editors, *Proc. of 4th Workshop on Deception, Fraud and Trust*, 2001.
8. R. Falcone and C. Castelfranchi. Trust dynamics: How trust is influenced by direct experiences and by trust itself. In *Proceedings of the 3rd International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-04)*, pages 740–747. New York, ACM, 2004.
9. K. Giffin. The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. *Psychological Bulletin*, 62:104–120, 1967.
10. A.J.I. Jones. On the concept of trust. *Decision Support Systems*, 33, 2002.
11. A.J.I. Jones and B.S. Firozabadi. On the characterisation of a trusting agent. Aspects of a formal approach. In C. Castelfranchi and Y-H. Tan, editors, *Trust and Deception in Virtual Societies*. Kluwer Academic Publisher, 2001.
12. R. Kohlas, J. Jonczyk, and R. Haenni. A trust evaluation method based on logic and probability theory. In Y. Karabulut, J. Mitchell, P. Herrmann, and C. D. Jensen, editors, *IFIPTM'08, 2nd Joint iTrust and PST Conferences on Privacy Trust Management and Security*, volume II of *Trust Management*, pages 17–32, Trondheim, Norway.
13. N. Laverny. *Raisonnement sur les actions et les observations, et programmes à base de croyances graduelles*. PhD thesis, Univeristé Paul Sabatier, Toulouse, 2006.
14. N. Laverny and J. Lang. From knowledge-based programs to graded belief-based programs part ii: Off-line reasoning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2005.
15. E. Lorini and R. Demolombe. From binary trust to graded trust in information sources: a logical perspective. In R. Falcone, S. Barber, J. Sabater-Mir, and M. Singh, editors, *Proceedings of the Workshop Trust in Agent Societies*. Springer, To appear.
16. T. Rea. Engending trust in electronic environments - roles of a trusted third party. In C. Castelfranchi and Y-H. Tan, editors, *Trust and Deception in Virtual Societies*. Kluwer Academic Publisher, 2001.
17. W. Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In W. L. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change and Statistics*, pages 105–134. Springer, 1988.