



Proceedings of the
Third International Workshop on
Formal Methods for Interactive Systems
(FMIS 2009)

A Logical Framework for Trust-Related Emotions

Jean-François Bonnefon, Dominique Longin and Manh-Hung Nguyen

15 pages

A Logical Framework for Trust-Related Emotions

Jean-François Bonnefon¹, Dominique Longin² and Manh-Hung Nguyen³

¹ bonnefon@univ-tlse2.fr

University of Toulouse, CNRS, CLLE, France

² Dominique.Longin@irit.fr

University of Toulouse, CNRS, IRIT, France

³ Manh-Hung.Nguyen@irit.fr

University of Toulouse, UPS, IRIT, France

Abstract: Emotion and trust are two important concerns for the elaboration of interaction systems that would be closer and more attractive to their users, in particular by endowing machines with the ability to predict, understand, and process emotions and trust. This paper attempts to construct a common logical framework for the representation of emotion and trust. This logical framework combines a logic of belief and choice, a logic of time, and a dynamic logic. Using this common framework, we identify formal relations between trust and emotions, for which we also provide behavioral validation.

Keywords: Modal logic, emotions, trust, distrust

1 Introduction

The rapidly growing field of affective computing aims at developing interaction systems that are closer and more attractive to their users, in particular by endowing machines with the ability to predict, understand, and process emotions (on the one hand), and trust (on the other hand). In this article, we introduce a unified logical approach to represent the cognitive structure of some emotions, of trust/distrust, and their relations at a formal level.

We formalize the concepts of emotions as well as trust/distrust based on cognitive models proposed by cognitive psychologists. Regarding emotions, we draw on cognitive theories (for more detail, see [SSJ01]) which assume that emotions are closely tied to changes in beliefs and desires. We capitalize on psychological models that allow to recognize and distinguish emotions based on their decomposition in cognitive factors particularly the cognitive structure of emotion of Ortony et al. [OCC88], the cognitive patterns of emotion of Lazarus [Laz91] and the belief-desire theory of emotion (BDTE) [Rei09, Dre95]. Similarly, we attempt to adhere closely to cognitive definition of trust [CF01] and distrust [CFL08].

Although there are tight conceptual connections between emotion and trust [Lah01], and although there were some separated formalization of the concepts of trust as the works of Herzig et al. [HLH⁺08], and the concepts of emotions such as the works of Adam et al. [AHL09] and Steunebrink et al. [SDM07b, SDM07a], there is not yet a common logic to represent them both. Our work aims at filling that gap by formally representing trust and emotions in a common logic;

this common logic will enable us to lay bare the formal relations between trust and emotion. The logic we offer is a combination of the logic of beliefs and choices as the one of Herzig and Longin [HL04] (a refinement from Cohen and Levesque [CL90]), the logic of time (introduced by Arthur Prior [Pri57]), and dynamic logic introduced by Fischer and Ladner [FL79] and Harel et al. [HKT00].

This paper is organized as follows: Part 2 introduces the logical framework. Part 3 formalizes the cognitive structure of some emotions, Part 4 formalizes the cognitive structure of trust and distrust. Part 5 shows some formal relations in the effect of trust/distrust on the emotions, and provides behavioral validation for these relations.

2 Logical Framework

Syntax. The syntactic primitives of our logic are as follows: a nonempty finite set of agents $AGT = \{i_1, i_2, \dots, i_n\}$, a nonempty finite set of atomic events $EVT = \{e_1, e_2, \dots, e_p\}$, and a nonempty set of atomic propositions $ATM = \{p_1, p_2, \dots\}$. The variables i, j, k, \dots denote agents. The expression $i_1:e_1 \in AGT \times EVT$ denotes an event e_1 intentionally caused by agent i_1 and e_1 is thus called an “action”. The variables α, β, \dots denote such actions. The language of our logic is defined by the following BNF :

$$\varphi := p \mid i:\alpha\text{-happens} \mid \neg\varphi \mid \varphi \vee \varphi \mid X\varphi \mid X^{-1}\varphi \mid G\varphi \mid Bel_i\varphi \mid Choice_i\varphi \mid Grd_I\varphi$$

where p ranges over ATM , $i:\alpha$ ranges over $AGT \times EVT$, $i:\alpha\text{-happens}$ ranges over ATM for each $i:\alpha \in AGT \times EVT$, and $I \subseteq AGT$. The classical boolean connectives \wedge (conjunction), \rightarrow (material implication), \leftrightarrow (material equivalence), \top (tautology) and \perp (contradiction) are defined from \neg (negation) and \vee (disjunction).

$i:\alpha\text{-happens}$ reads “agent i is just about to perform the action α ”; $X\varphi$ reads “ φ will be true next instant”; $X^{-1}\varphi$ reads “ φ was true at the previous instant”; $G\varphi$ reads “henceforth, φ is true”; $Bel_i\varphi$ reads “agent i believes that φ is true”; $Choice_i\varphi$ reads “agent i prefers that φ be true”; $Grd_I\varphi$ reads “ φ is publicly grounded between the agents in group I ” (It is nothing else than a standard common belief operator). We define the following abbreviations:

$$\begin{aligned} i:\alpha\text{-done} &\stackrel{def}{=} X^{-1}i:\alpha\text{-happens} && (\text{Def}_{i:\alpha\text{-done}}) \\ \text{Happens}_{i:\alpha}\varphi &\stackrel{def}{=} i:\alpha\text{-happens} \wedge X\varphi && (\text{Def}_{\text{Happens}_{i:\alpha}}) \\ \text{After}_{i:\alpha}\varphi &\stackrel{def}{=} i:\alpha\text{-happens} \rightarrow X\varphi && (\text{Def}_{\text{After}_{i:\alpha}}) \\ \text{Done}_{i:\alpha}\varphi &\stackrel{def}{=} i:\alpha\text{-done} \wedge X^{-1}\varphi && (\text{Def}_{\text{Done}_{i:\alpha}}) \end{aligned}$$

$$\begin{aligned}
 F\varphi &\stackrel{def}{=} \neg G\neg\varphi && (\text{Def}_F) \\
 \text{Goal}_i\varphi &\stackrel{def}{=} \text{Choice}_i\text{FBel}_i\varphi && (\text{Def}_{\text{Goal}_i}) \\
 \text{Intend}_i\alpha &\stackrel{def}{=} \text{Choice}_i\text{Fi}\alpha\text{-happens} && (\text{Def}_{\text{Intend}_i}) \\
 \text{Capable}_i\alpha &\stackrel{def}{=} \neg\text{After}_{i:\alpha}\perp && (\text{Def}_{\text{Capable}_i}) \\
 \text{Possible}_i\varphi &\stackrel{def}{=} \neg\text{Bel}_i\neg\varphi && (\text{Def}_{\text{Possible}_i}) \\
 \text{Awareness}_i\varphi &\stackrel{def}{=} X^{-1}\neg\text{Bel}_i\varphi \wedge \text{Bel}_i\varphi && (\text{Def}_{\text{Awareness}_i})
 \end{aligned}$$

$i:\alpha$ -done reads “agent i has done action α ”; $\text{Happens}_{i:\alpha}\varphi$ reads “agent i is doing action α and φ will be true next instant”; $\text{After}_{i:\alpha}\varphi$ reads “ φ is true after any execution of α by i ”; $\text{Done}_{i:\alpha}\varphi$ reads “agent i has done action α and φ was true at previous instant”; $F\varphi$ reads “ φ will be true in some future instants”; $\text{Goal}_i\varphi$ reads “agent i has the goal (chosen preference) that φ be true”; $\text{Intend}_i\alpha$ reads “agent i intends to do α ”; $\text{Capable}_i\alpha$ reads “agent i is capable to do α ”; $\text{Possible}_i\varphi$ reads “agent i believes that it is possible φ ”; $\text{Awareness}_i\varphi$ reads “agent i has just experienced that φ is true”.

Semantics. For temporal operators, we use a semantics based on linear time described by a sequence (or story) of time points. (This semantics is very close to CTL* [CES86]) A frame \mathcal{F} is a 4-tuple $\langle H, \mathcal{B}, \mathcal{C}, \mathcal{G} \rangle$ where: H is a set of stories that are represented as sequences of time points, where each time point is identified by an integer $z \in \mathbb{Z}$, a time point z in a story h is called a situation $\langle h, z \rangle$; \mathcal{B} is the set of all \mathcal{B}_i such that $\mathcal{B}_i(h, z)$ denotes the set of stories believed as being possible by the agent i in the situation $\langle h, z \rangle$; \mathcal{C} is the set of all \mathcal{C}_i such that $\mathcal{C}_i(h, z)$ denotes the set of stories chosen by the agent i in the situation $\langle h, z \rangle$; \mathcal{G} is the set of all \mathcal{G}_I such that $\mathcal{G}_I(h, z)$ denotes the set of stories which are publicly grounded in the group I of agents, in the situation $\langle h, z \rangle$.

All the accessibility relations \mathcal{B}_i are serial¹, transitive² and euclidean³. This semantic is completely standard in epistemic logic (see [Hin62, GG06]) All the accessibility relations \mathcal{G}_I are serial, transitive and euclidean (This is similar to the operator group grounding introduced by Gaudou et al. [GHL06]). All the accessibility \mathcal{C}_i are serial. Moreover, we impose for every $z \in \mathbb{Z}$ that: if $h' \in \mathcal{B}_i(h, z)$ then $\mathcal{C}_i(h, z) = \mathcal{C}_i(h', z)$. It means that if an agent believes that the world h' is possible from the world h , then the set of his/her preference worlds from h and h' are the same. In other terms, the worlds an agent prefers and the ones that agent believes that s/he prefers are the same (briefly, the agent is conscious about his/her preferences, and s/he prefers what s/he believes that s/he prefers).

A model \mathcal{M} is a couple $\langle \mathcal{F}, \mathcal{V} \rangle$ where \mathcal{F} is a frame and \mathcal{V} is a function associating each atomic proposition p with the set $\mathcal{V}(p)$ of couple (h, z) where p is true. Truth conditions are

¹ for every $w_1 \in \mathcal{M}$, there is w_2 such that $w_1 \mathcal{B}_i w_2$

² if $w_1 \mathcal{B}_i w_2$ and $w_2 \mathcal{B}_i w_3$, then $w_1 \mathcal{B}_i w_3$

³ if $w_1 \mathcal{B}_i w_2$ and $w_1 \mathcal{B}_i w_3$, then $w_2 \mathcal{B}_i w_3$

defined as follows:

$$\begin{aligned}
\mathcal{M}, h, z \models p &\text{ iff } (h, z) \in \mathcal{V}(p) \\
\mathcal{M}, h, z \models X\varphi &\text{ iff } \mathcal{M}, h, z+1 \models \varphi \\
\mathcal{M}, h, z \models X^{-1}\varphi &\text{ iff } \mathcal{M}, h, z-1 \models \varphi \\
\mathcal{M}, h, z \models G\varphi &\text{ iff } \mathcal{M}, h, z' \models \varphi \text{ for every } z' \geq z \\
\mathcal{M}, h, z \models \text{Bel}_i \varphi &\text{ iff } \mathcal{M}, h', z \models \varphi \text{ for every } (h', z) \in \mathcal{B}_i(h, z) \\
\mathcal{M}, h, z \models \text{Choice}_i \varphi &\text{ iff } \mathcal{M}, h', z \models \varphi \text{ for every } (h', z) \in \mathcal{C}_i(h, z) \\
\mathcal{M}, h, z \models \text{Grd}_I \varphi &\text{ iff } \mathcal{M}, h', z \models \varphi \text{ for every } (h', z) \in \mathcal{G}_I(h, z)
\end{aligned}$$

Other truth conditions are defined as usual.

Axiomatics. Due to our linear time semantics, the temporal operators satisfy the following principles:

$$\begin{aligned}
i:\alpha\text{-happens} &\leftrightarrow Xi:\alpha\text{-done} & (1) \\
X\varphi &\leftrightarrow \neg X\neg\varphi & (2) \\
\varphi &\leftrightarrow XX^{-1}\varphi & (3) \\
\varphi &\leftrightarrow X^{-1}X\varphi & (4) \\
G\varphi &\leftrightarrow \varphi \wedge XG\varphi & (5) \\
G(\varphi \rightarrow X\varphi) &\rightarrow (\varphi \rightarrow G\varphi) & (6)
\end{aligned}$$

Bel_i and Choice_i operators are defined in a normal modal logic plus (D) axioms. Thus, if \Box represents a Bel_i operator or Choice_i operator:

$$\begin{aligned}
&\frac{\varphi}{\Box\varphi} & (\text{RN}_{\Box}) \\
\Box(\varphi \rightarrow \psi) &\rightarrow (\Box\varphi \rightarrow \Box\psi) & (\text{K}_{\Box}) \\
\Box\varphi &\rightarrow \neg\Box\neg\varphi & (\text{D}_{\Box})
\end{aligned}$$

For example, axiom D_{\Box} applied to operator Bel_i is D_{Bel_i} , which is described as: $\text{Bel}_i \varphi \rightarrow \neg\text{Bel}_i \neg\varphi$.

(RN_{\Box}) means that all theorems are believed (respectively: chosen) by every agent i ; (K_{\Box}) means that beliefs (respectively: choices) are closed under material implication for every agent i ; (D_{\Box}) means that beliefs (respectively: choices) of every agent i are rational: they cannot be contradictory.

The Bel_i operators satisfy the following principles of introspection:

$$\begin{aligned}
\text{Bel}_i \varphi &\leftrightarrow \text{Bel}_i \text{Bel}_i \varphi & (4_{\text{Bel}_i}) \\
\neg\text{Bel}_i \varphi &\leftrightarrow \text{Bel}_i \neg\text{Bel}_i \varphi & (5_{\text{Bel}_i})
\end{aligned}$$

that means that agent i is conscious of its beliefs and of its disbeliefs.

The following principle follows from the semantical constraint between belief accessibility relation and choice accessibility relation, and from axiom (D_{\Box}) for Bel_i :

$$\text{Choice}_i \varphi \leftrightarrow \text{Bel}_i \text{Choice}_i \varphi \quad (4_{BC})$$

$$\neg \text{Choice}_i \varphi \leftrightarrow \text{Bel}_i \neg \text{Choice}_i \varphi \quad (5_{BC})$$

that means that agent i is conscious of its choices and of its dischoices.

The sound and complete axiomatization of Grd_I operator is defined as the one of common belief operator (also called mutual belief), which is closed to the operator described in Walton and Krabbe [WK95], also introduced by Gaudou et al. [GHL06]:

$$\frac{\varphi}{\text{Grd}_I \varphi} \quad (\text{RN}_{\text{Grd}_I})$$

$$\text{Grd}_I(\varphi \rightarrow \psi) \rightarrow (\text{Grd}_I \varphi \rightarrow \text{Grd}_I \psi) \quad (\text{K}_{\text{Grd}_I})$$

$$\text{Grd}_I \varphi \rightarrow \neg \text{Grd}_I \neg \varphi \quad (\text{D}_{\text{Grd}_I})$$

$$\text{Grd}_I \varphi \rightarrow \text{Grd}_I \text{Grd}_I \varphi \quad (4_{\text{Grd}_I})$$

$$\neg \text{Grd}_I \varphi \rightarrow \text{Grd}_I \neg \text{Grd}_I \varphi \quad (5_{\text{Grd}_I})$$

Axiom $(\text{RN}_{\text{Grd}_I})$ means that every tautology is public ground. Axiom $(\text{K}_{\text{Grd}_I})$ means that if φ is publicly grounded in I and that φ implies ψ then ψ is also publicly grounded in I . Axiom $(\text{D}_{\text{Grd}_I})$ means that the set of grounded informations is consistent: it can not be the case that both φ and $\neg \varphi$ are simultaneously grounded. The positive introspection axiom (4_{Grd_I}) and negative introspection axiom (5_{Grd_I}) account for the public character of Grd_I . From these collective awareness results: if φ has (resp. has not) been grounded then it is established that φ has (resp. has not) been grounded.

Linear time semantics entail the following principles:

$$\text{G}\varphi \rightarrow \text{After}_{i:\alpha} \varphi \quad (7)$$

$$\text{Happens}_{i:\alpha} \varphi \rightarrow \text{After}_{j:\beta} \varphi \quad (8)$$

$$\text{After}_{i:\alpha} \varphi \leftrightarrow \neg \text{Happens}_{i:\alpha} \neg \varphi \quad (9)$$

Axiom (7) describe the relationship between time and action: if henceforth φ is true then after every action α of every agent i , φ will be true. (Note that the converse is not valid: it is possible that φ be true after every action α of every agent i performed in a situation $\langle h, z \rangle$, and that φ be false at time $z' > z$.)

As time is linear, actions are deterministic on a given history. Thus, axiom (8) reads: if agent i is just about to perform α after what φ will be true, then after every performance of every action β by every agent j , φ will be true. In other words, if action α leads to a time point where φ is true, then every action performed by every agent leads to this time point.

Finally, axiom (9) means that $\text{After}_{i:\alpha}$ and $\text{Happens}_{i:\alpha}$ operators are dual operators. This property is fair with respect to dynamic logic [HKT00].

3 Formalization of the cognitive structure of emotion

In this section, we present the formalization of emotions, based on their cognitive structure as proposed by Ortony et al. [OCC88], Frijda [Fri86] as well as those of Reisenzei [Rei09] and Scherer et al. [Sch01].

Joy/Distress. The cognitive structure of *Joy* consists of two main factors: (i) a proposition φ is desirable for agent i , and (ii) agent i just experienced that φ is the case. To formalize the first factor, we consider that agent i desiring φ means that i wants φ to be the case. So we formalize desire as a goal (chosen preference). Therefore, the first factor is potentially formalized as $\text{Goal}_i \varphi$, the second factor may be formalized as $\text{Bel}_i \varphi$.

However, we assume that emotion is triggered at the moment when all its factors are fulfilled, and that its intensity then decreases with time [dS01, Fri86]. Accordingly, we include a time factor into most emotional formulas. Thus, the first factor of *Joy* in particular means that agent i now recalls that at the previous instant, s/he desired φ , until experiencing that φ was in fact true: $\text{Bel}_i X^{-1} \text{Goal}_i \varphi$. It means that in order to be joyful, agent i must keep in mind his desire in the previous instant. Hereafter, we add this analysis for almost emotional formulas. The second factor means that agent i has just experienced that φ is true and did not previously know it: $\text{Awareness}_i \varphi$.

The same analysis applies to *Distress*, except that in the first factor of *Distress*, φ is undesirable for agent i , which we assume to mean that agent i desired $\neg\varphi$: $\text{Bel}_i X^{-1} \text{Goal}_i \neg\varphi$. We accordingly formalize the concept of *Joy* and *Distress*:

Definition 1 (Joy/Distress)

$$\begin{aligned} \text{Joy}_i \varphi &\stackrel{\text{def}}{=} \text{Bel}_i X^{-1} \text{Goal}_i \varphi \wedge \text{Awareness}_i \varphi \\ \text{Distress}_i \varphi &\stackrel{\text{def}}{=} \text{Bel}_i X^{-1} \text{Goal}_i \neg\varphi \wedge \text{Awareness}_i \varphi \end{aligned}$$

To illustrate the definition of *Joy*, we can say that an individual is joyful when he has just realized that he won the lottery ($\text{Awareness}_{\text{man}}(\text{win lottery})$) with the trivial assumption that he had been desiring to win the lottery ($\text{Bel}_{\text{man}} X^{-1} \text{Goal}_{\text{man}}(\text{win lottery})$). In contrast, to illustrate the definition of *Distress*, we can say that an individual feels distress when she learns she has lost her job ($\text{Awareness}_{\text{woman}}(\text{lost job})$) assuming that she had the goal not to lose her job ($\text{Bel}_{\text{woman}} X^{-1} \text{Goal}_{\text{woman}} \neg(\text{lost job})$).

Hope/Fear. The cognitive structure of *Hope* consists of two factors: (i) a proposition φ is desirable for agent i , and (ii) agent i believes that φ may be true in the future. To formalize the first factor, we consider that φ is not true at the moment when i hopes for it: $\text{Goal}_i \varphi$.

We interpret the second factor, as meaning that among all of possible future worlds, agent i believes that there is at least one world in which φ will be the case. In other terms, agent i does not believe that φ will be false in all of possible future worlds: $\text{Possible}_i F\varphi$. If i believes that φ can never be the case in all of possible future worlds, then i has no ground for hope.

The same analysis applies to *Fear*, except that φ is now undesirable for agent i : $\text{Goal}_i \neg\varphi$. We accordingly formalize the concept of *Hope* and *Fear*:

Definition 2 (Hope/Fear)

$$\begin{aligned} \text{Hope}_i \varphi &\stackrel{\text{def}}{=} \text{Goal}_i \varphi \wedge \text{Possible}_i F\varphi \\ \text{Fear}_i \varphi &\stackrel{\text{def}}{=} \text{Goal}_i \neg\varphi \wedge \text{Possible}_i F\varphi \end{aligned}$$

For example, a debutante is hopeful about being asked to dance, for she thinks it is possible ($\text{Possible}_{\text{girl}}F(\text{being asked to dance})$) and this is what she wants ($\text{Goal}_{\text{girl}}(\text{being asked to dance})$). In contrast, an employee fears to be fired when he does not wish to be fired ($\text{Goal}_{\text{employee}}\neg(\text{fired})$) but believes it is a possibility $\text{Possible}_{\text{employee}}F(\text{to be fired})$.

Satisfaction/Disappointment. The cognitive structure of *Satisfaction* consists of three factors: (i) agent i desired a proposition φ , (ii) agent i used to believe that φ might be true in the near future, and (iii) agent i now experiences that φ is really the case. The first two factors mean that now, agent keeps in mind that at the previous instant, s/he desired φ and believed that φ could be true in the future ($\text{Bel}_iX^{-1}(\text{Goal}_i\varphi \wedge \text{Possible}_iF\varphi)$) (cf. the analysis of the second factor of *Hope*). The last factor means that i now experiences that φ is true, but did not know it the previous instant ($\text{Awareness}_i\varphi$).

The difference in the case of *Disappointment* is agent recalls that, in the previous instant, s/he desired $\neg\varphi$ instead of φ , and s/he believed that $\neg\varphi$ was possibly true in the future ($\text{Bel}_iX^{-1}(\text{Goal}_i\neg\varphi \wedge \text{Possible}_iF\neg\varphi)$). We formalize *Satisfaction* and *Disappointment* as

Definition 3 (Satisfaction/Disappointment)

$$\text{Satisfaction}_i\varphi \stackrel{\text{def}}{=} \text{Bel}_iX^{-1}(\text{Goal}_i\varphi \wedge \text{Possible}_iF\varphi) \wedge \text{Awareness}_i\varphi$$

$$\text{Disappointment}_i\varphi \stackrel{\text{def}}{=} \text{Bel}_iX^{-1}(\text{Goal}_i\neg\varphi \wedge \text{Possible}_iF\neg\varphi) \wedge \text{Awareness}_i\varphi$$

For example, when the debutante realizes that she is indeed asked to dance ($\text{Awareness}_{\text{girl}}(\text{asked to dance})$) she is satisfied. Were she not to be asked to dance ($\text{Awareness}_{\text{girl}}(\text{not asked to dance})$), she would feel disappointed.

We can point out the relations between *Satisfaction*, *Disappointment* and *Hope*:

$$\text{Satisfaction}_i\varphi \leftrightarrow \text{Bel}_iX^{-1}\text{Hope}_i\varphi \wedge \text{Awareness}_i\varphi \quad (10)$$

$$\text{Disappointment}_i\varphi \leftrightarrow \text{Bel}_iX^{-1}\text{Hope}_i\neg\varphi \wedge \text{Awareness}_i\varphi \quad (11)$$

The relation between *Satisfaction* and *Joy* can be formalized as Proposition 1: if we feel satisfaction about something, then we will also feel joy about it.

Proposition 1 (*Satisfaction implies Joy*)

$$\text{Satisfaction}_i\varphi \rightarrow \text{Joy}_i\varphi$$

Fear-confirmed/Relief. The cognitive structure of *Fear-confirmed* consists of three factors: (i) a proposition φ was undesirable for agent i , (ii) agent i believed that φ might be true in the near future, and (iii) agent i now experiences that φ is really true.

We use the same analysis as for *Satisfaction*, except agent recalls that in the previous instant, $\neg\varphi$ was desirable for agent i ($\text{Bel}_iX^{-1}\text{Goal}_i\neg\varphi$).

The difference in the case of *Relief* is agent recalls that, in the previous instant, s/he desired φ ($\text{Bel}_iX^{-1}\text{Goal}_i\varphi$), and believed that $\neg\varphi$ might be true in the near future ($\text{Bel}_iX^{-1}(\text{Goal}_i\varphi \wedge \text{Possible}_iF\neg\varphi)$). We formalize *Fear-confirmed* and *Relief* as:

Definition 4 (Fear-confirmed/Relief)

$$\text{FearConfirmed}_i \varphi \stackrel{\text{def}}{=} \text{Bel}_i X^{-1} (\text{Goal}_i \neg \varphi \wedge \text{Possible}_i F \varphi) \wedge \text{Awareness}_i \varphi$$

$$\text{Relief}_i \varphi \stackrel{\text{def}}{=} \text{Bel}_i X^{-1} (\text{Goal}_i \varphi \wedge \text{Possible}_i F \neg \varphi) \wedge \text{Awareness}_i \varphi$$

For example, the employee's fear of being fired is confirmed when he learns that he is indeed about to be fired ($\text{Awareness}_{\text{employee}}(\text{fired})$) which he had been afraid of ($\text{Bel}_{\text{employee}} X^{-1} (\text{Goal}_{\text{employee}} \neg(\text{fired}) \wedge \text{Possible}_{\text{employee}} F(\text{fired}))$). In contrast, were he to learn that he is not going to be fired ($\text{Awareness}_{\text{employee}}(\text{not fired})$), he would feel relief.

We can also point out the relations between *Fear-confirmed*, *Relief* and *Fear*:

$$\text{FearConfirmed}_i \varphi \leftrightarrow \text{Bel}_i X^{-1} \text{Fear}_i \varphi \wedge \text{Awareness}_i \varphi \quad (12)$$

$$\text{Relief}_i \varphi \leftrightarrow \text{Bel}_i X^{-1} \text{Fear}_i \neg \varphi \wedge \text{Awareness}_i \varphi \quad (13)$$

The relation between *Fear-confirmed* and *Distress* is stated in Proposition 2: if our fears about something are confirmed, then we feel distressed.

Proposition 2 (*Fear-confirmed implies Distress*)

$$\text{FearConfirmed}_i \varphi \rightarrow \text{Distress}_i \varphi$$

4 Formalization of Trust

We now present the formalization of trust and distrust based on the cognitive definition of Castelfranchi and colleagues [CF01, CFL08].

Trust. We formalize the concept of trust based on Castelfranchi and Falcone's definition [CF01] of trust in action which says that agent i trusts agent j to ensure φ by performing action α if and only if agent i desires to achieve φ ($\text{Goal}_i \varphi$), and agent i expects that: (i) φ can be achieved by doing action α ($\text{Bel}_i \text{After}_{j:\alpha} \varphi$); (ii) agent j is able to perform action α ($\text{Bel}_i \text{Capable}_j \alpha$); and (iii) agent j has the intention to do such an action ($\text{Bel}_i \text{Intend}_j \alpha$).

However, these three factors are only necessary conditions, but not sufficient ones. For example, imagine that a robber wants to steal something located on the second floor of a mansion. There is a nurse on the first floor. The robber desires that the nurse stays where she is, because it makes his robbery possible. He also believes that it is possible that the nurse will stay where she is, and that it is actually her intention. Thus, the three conditions are satisfied, but we are reluctant nonetheless to say that the robber trusts the nurse to stay where she is in order to allow for his stealing, because there is no agreement between the nurse (trustee) and the robber (trustor). So here we need to add another condition for trust: an agreement between trustor and trustee that the trustee will perform such an action ($\text{Grd}_I \text{trustee} : \alpha\text{-happens}$), where $I = \{\text{trustor}, \text{trustee}\}$. We accordingly formalize the concept of trust as:

Definition 5 (Trust)

$$\text{Trust}_{i,j}(\alpha, \varphi) \stackrel{\text{def}}{=} \text{Goal}_i \varphi \wedge \text{Bel}_i \text{After}_{j:\alpha} \varphi \wedge \text{Bel}_i \text{Capable}_j \alpha \wedge \text{Bel}_i \text{Intend}_j \alpha \wedge \text{Grd}_{\{i,j\}} j:\alpha\text{-happens}$$

For example, a boss trusts his secretary to prepare a report in order to present it at a company meeting because the boss desires the report ($\text{Goal}_{\text{boss}}(\text{report})$), and in his opinion, the report can be possibly ready after the secretary prepares it ($\text{Bel}_{\text{boss}} \text{After}_{\text{secretary:prepare}}(\text{report})$), the secretary has the ability and intention to prepare the report ($\text{Bel}_{\text{boss}} \text{Capable}_{\text{secretary}}(\text{prepare}) \wedge \text{Bel}_{\text{boss}} \text{Intend}_{\text{secretary}}(\text{prepare})$). It is clear that in the relation between the boss and his secretary, there is an agreement that the secretary will prepare the report in time ($\text{Grd}_{\text{boss,secretary}} \text{secretary} : \text{prepare-happens}$).

Distrust. We also adopt the definition of distrust given by Castelfranchi et al. [CFL08] which says that agent i distrusts agent j to ensure φ by performing action α if and only if agent i desires to achieve φ ($\text{Goal}_i \varphi$), and agent i believes that at least one of these conditions is fulfilled: (i) agent j is not in the capacity to do action α : $\text{Bel}_i \neg \text{After}_{j:\alpha} \varphi$, or (ii) agent j is able to do α but he has not intention to do α : $\text{Possible}_i \text{After}_{j:\alpha} \varphi \wedge \text{Bel}_i \neg \text{Intend}_j \alpha$. We accordingly formalize this concept as:

Definition 6 (Distrust)

$$\text{DisTrust}_{i,j}(\alpha, \varphi) \stackrel{\text{def}}{=} \text{Goal}_i \varphi \wedge (\text{Bel}_i \neg \text{After}_{j:\alpha} \varphi \vee (\text{Possible}_i \text{After}_{j:\alpha} \varphi \wedge \text{Bel}_i \neg \text{Intend}_j \alpha))$$

For example, in spite of desiring the report ($\text{Goal}_{\text{boss}}(\text{report})$), the boss does not trust a new employee to prepare it because he believes the new employee is unable to perform that task ($\text{Bel}_{\text{boss}} \neg \text{After}_{\text{employee:prepare}}(\text{report})$).

From this definition, we can decompose the concept of distrust based only on the ability of trustee:

Definition 7 (Distrust based on ability)

$$\text{C-DisTrust}_{i,j}(\alpha, \varphi) \stackrel{\text{def}}{=} \text{Goal}_i \varphi \wedge \text{Bel}_i \neg \text{After}_{j:\alpha} \varphi$$

5 Trust-Related Emotions

5.1 Formal Relations

Trust and Hope. *Trust* and *Hope* have an important relation because they both feature a positive expectation [CF01]. When i trusts j , i has a positive expectation about j 's power and performance. *Hope* also implies some positive expectation. The greater the expectations, the deeper the trust; and, conversely, the deeper the disappointment when expectations are unrealized [Bry07]. We formalize the former relation as Proposition 3, the latter as Proposition 5.

Proposition 3 (*Trust implies Hope*)

$$\text{Trust}_{i,j}(\alpha, \varphi) \rightarrow \text{Hope}_i \varphi$$

This means that when we trust someone about an action that will bring some results, we are hopeful that the results will be obtained. For example, in a commercial transaction, when the buyer trusts his seller to send him a product after payment ($\text{Trust}_{\text{buyer},\text{seller}}(\text{send}, \text{receipt})$), he will be hopeful that he will receive the product ($\text{Hope}_{\text{buyer}} \text{receive product}$). This proposition will be proved by applying Lemma 1: if we believe that φ is true after every execution of action α , and that someone is able to do α , then we believe that there is at least a future world in which φ is true.

Lemma 1

$$\text{Bel}_i \text{After}_{j:\alpha} \varphi \wedge \text{Bel}_i \text{Capable}_j \alpha \rightarrow \text{Possible}_i \text{F} \varphi$$

Once we trust someone to do an action to bring us something, we hope for the positive result of the action. In case of success, we feel satisfaction (formalized as Proposition 4). Conversely, in case of failure, we feel disappointment (formalized as Proposition 5).

Proposition 4 (*Successful Trust implies Satisfaction*)

$$\text{Bel}_i \text{Done}_{j:\alpha} \text{Trust}_{i,j}(\alpha, \varphi) \wedge \text{Awareness}_i \varphi \rightarrow \text{Satisfaction}_i \varphi$$

This means that when we believe that what we trusted has now occurred, we are satisfied about it. For example, when the boss trusted his secretary to prepare the report ($\text{Done}_{\text{secretary}:\text{prepare}} \text{Trust}_{\text{boss},\text{secretary}}(\text{prepare}, \text{having report})$), and on the morning of the day after, he has received the report ($\text{Bel}_{\text{boss}} \text{having report}$), then he is satisfied ($\text{Satisfaction}_{\text{boss}} \text{having report}$). This proposition has a corollary which is deduced from Proposition 1 and 4: When we experience that what we trusted has really occurred, we will also feel joy about it.

Corollary 1

$$\text{Bel}_i \text{Done}_{j:\alpha} \text{Trust}_{i,j}(\alpha, \varphi) \wedge \text{Awareness}_i \varphi \rightarrow \text{Joy}_i \varphi$$

Proposition 5 (*Unsuccessful Trust implies Disappointment*)

$$\text{Bel}_i \text{Done}_{j:\alpha} \text{Trust}_{i,j}(\alpha, \varphi) \wedge \text{Awareness}_i \neg \varphi \rightarrow \text{Disappointment}_i \neg \varphi$$

This means that we feel disappointed if what we trusted does not in fact occur. For example, a businessman trusted his partner to arrive on time to negotiate a contract. The businessman feels disappointed if the partner has not yet arrived at the scheduled time.

DisTrust and Fear. Distrust features a negative expectation, involving fear of the other [LW00, AACSO8]. We state the relation between *Distrust* based on ability and *Fear* as Proposition 6.

Proposition 6 (*DisTrust implies Fear*)

$$C\text{-DisTrust}_{i,j}(\alpha, \varphi) \rightarrow \text{Fear}_i \neg \varphi$$

This means that if we distrust someone to do an action to bring us something then we fear that our desire might not be fulfilled. For example, the boss might distrust his assistant with the preparation of a report he needs, and more specifically distrusts him to finish the report by the next morning ($\text{DisTrust}_{\text{boss,assistant}}(\text{finish}, \text{report})$). Therefore, he is fearful that he might miss the report the next morning ($\text{Fear}_{\text{boss}} \neg \text{report}$). This proposition will be proved by applying Lemma 2: if we believe that someone is unable to do an action to bring about something, then we believe that there is at least a future world without the expected result of this action.

Lemma 2

$$\text{Bel}_i \neg \text{After}_{j:\alpha} \varphi \rightarrow \text{Possible}_i F \neg \varphi$$

Once we distrust someone to do an action to bring about something, we experience fear. If the results are indeed negative, we feel fear-confirmed (formalized as Proposition 7). If, however the action is in fact successfully performed, we feel relief (formalized as Proposition 8).

Proposition 7 (*Confirmation of DisTrust implies Fear-confirmed*)

$$\text{Bel}_i \text{Done}_{j:\alpha} C\text{-DisTrust}_{i,j}(\alpha, \varphi) \wedge \text{Awareness}_i \neg \varphi \rightarrow \text{FearConfirmed}_i \neg \varphi$$

If the boss realizes that his assistant really did not finish the report ($\text{Bel}_{\text{boss}} \neg \text{report}$), he feels fear-confirmed ($\text{FearConfirmed}_{\text{boss}} \neg \text{report}$). Combining the two Propositions 2 and 7, we arrive at a corollary: when we experience that what we distrusted has now happened, we feel distressed about it.

Corollary 2

$$\text{Bel}_i \text{Done}_{j:\alpha} C\text{-DisTrust}_{i,j}(\alpha, \varphi) \wedge \text{Awareness}_i \neg \varphi \rightarrow \text{Distress}_i \neg \varphi$$

Proposition 8 (*Non-confirmation of DisTrust implies Relief*)

$$\text{Bel}_i \text{Done}_{j:\alpha} C\text{-DisTrust}_{i,j}(\alpha, \varphi) \wedge \text{Awareness}_i \varphi \rightarrow \text{Relief}_i \varphi$$

If the boss discovers that his assistant did in fact finish the report ($\text{Bel}_{\text{boss}} \text{report}$), he feels relieved ($\text{Relief}_{\text{boss}} \text{report}$).

5.2 Behavioral validation

Although the propositions that we proved in the previous section are intuitively plausible, some of them have not yet received behavioral validation from the field of experimental psychology. We decided to collect empirical data concerning three propositions in this article, related to the emotions that follow trust when it is confirmed (Proposition 4), and when it is unconfirmed (Proposition 5); and the emotions that follow distrust, when it is unconfirmed (Proposition 8)⁴.

⁴ We could not test Proposition 7 for a linguistic reason: Neither in French nor in Vietnamese (the two languages used in our experiment) could we find an everyday term equivalent to ‘fear confirmed’.

	Satisfaction		Relief		Disappointment	
	Trust	Distrust	Trust	Distrust	Trust	Distrust
Success	4.9 (1.5)	4.6 (1.6)	2.8 (1.9)	3.6 (1.9)	1.1 (0.6)	1.3 (0.8)
Failure	1.1 (0.5)	1.4 (1.0)	1.3 (1.0)	1.3 (0.9)	4.6 (1.7)	3.2 (1.4)

Table 1: Mean and standard deviations of affective ratings, as a function of Trust and Outcome.

Following the analysis in (Section 4) which argues that trust is the conjunction of the intention, the capacity, and the agreement of trustee, the presence of *Agreement* is intentionally fixed for the future test. We therefore operationalize *Trust* as the conjunction of *Intention* and *Capacity*, and *Distrust* as the three remaining cases. Participants to the survey read 8 different stories, following a $2 \times 2 \times 2$ within-subject design. The variables manipulated in the stories were *Intention* (Yes/No), *Capacity* (Yes/No), and *Outcome* (Success/Failure). As an example, here is the story corresponding to *Intention* = Yes, *Capacity* = Yes, and *Outcome* = Success.

Mr. Boss is the marketing director of a big company. He needs an important financial report before a meeting tomorrow morning, but he has no time to write it because of other priorities. He asks Mr. Support to prepare it and put it on his desk before tomorrow morning.

- Mr. Boss believes that Mr. Support has the intention to prepare the report in time.
- Mr. Boss believes that Mr. Support is able to prepare the report in time.

The morning after, Mr. Boss finds the report on his desk when he arrives. In your opinion, what does he feel?

In the condition *Intention* = No, “Mr. Boss believes that Mr. Support has the intention to prepare the report in time” was replaced with “Mr. Boss believes that Mr. Support has no intention to prepare the report in time.” In the condition *Capacity* = No, “Mr. Boss believes that Mr. Support is able to prepare the report in time” was replaced with “Mr. Boss believes that Mr. Support is unable to prepare the report in time.” Finally, in the condition *Outcome* = Failure, “Mr. Boss finds the report on his desk when he arrives” was replaced with “Mr. Boss does not find the report on his desk when he arrives.”

After reading each story, participants rated the extent to which the main character would feel each of 7 emotions, which included our target emotions, satisfaction, disappointment, and relief; but also some emotions that we included for exploratory purposes, such as anger or thankfulness. Ratings used a 6-point scale anchored at *Not at all* and *Totally*.

A total of 100 participants took part in an online survey. The survey was offered in two languages, French (30% of the final sample) and Vietnamese (70%). Language was entered as a control variable in all statistical analyses, but added only a small overall main effect on participants’ responses, and will not be discussed any further.

Descriptive statistics are displayed in Table 1. Participants’ responses were analyzed by means of a repeated-measure analysis of variance, aimed at detecting statistically reliable effects of Trust and Outcome on our emotions of interest.

Satisfaction. Unsurprisingly, the analysis of variance detected a huge effect of Outcome, $F(1, 98) = 597$, $p < .001$, accounting for most of the observed variance, $\eta_p^2 = .86$. In other terms,

Satisfaction is almost perfectly predicted by Outcome alone. The analysis, however, also detects a comparatively small interaction effect Outcome \times Trust, $F(1,98) = 8.8$, $p < .01$, $\eta_p^2 = .08$, reflecting the fact that success is even more pleasant in case of trust. Table 1 shows that the biggest score of *Satisfaction* is in the case of Trust follows a Success: $M = 4.9$, $SD < 1.5$. The data are in line with what was expected from Proposition 4.

Relief. The analysis detected main effects of Trust, $F(1,98) = 19.1$, $p < .001$, $\eta_p^2 = .23$; and Outcome, $F(1,98) = 127$, $p < .001$, $\eta_p^2 = .80$. However, these main effects were qualified by an interaction effect Trust \times Outcome, $F(1,98) = 12.3$, $p < .001$, $\eta_p^2 = .31$. Table 1 shows that the score of *Relief* is especially high in the case of Success is obtained despite of Distrust: $M = 3.6$, $SD < 1.9$. This interaction reflects our expectation (Proposition 8).

Disappointment. The analysis detected main effects of Trust, $F(1,98) = 28.4$, $p < .001$, $\eta_p^2 = .16$; and Outcome, $F(1,98) = 389$, $p < .001$, $\eta_p^2 = .56$. However, these main effects were qualified by an interaction effect Trust \times Outcome, $F(1,98) = 44.7$, $p < .001$, $\eta_p^2 = .11$. Table 1 shows that the score of *Disappointment* is especially high in the case of Failure is obtained despite of Trust: $M = 4.6$, $SD < 1.7$. This interaction reflects our expectation (Proposition 5).

6 Conclusion

This paper introduced a logical framework that can represent the cognitive structure of emotions, trust, and the formal relations between them. In other terms, it enables to represent the effect of trust (and distrust) on emotions. Furthermore, this logical framework respects the instantaneity of emotions that previous logics of emotions did not fulfill. Finally, the formal relations between emotion and trust laid bare by the logical framework were subjected to a behavioral validation following the methods of experimental psychology. The success of this behavioral validation gives strong support to our approach, which is shown to capture lay users' intuitions about trust-related emotion.

Although we have added time factor into almost emotional formulas, which enables to eliminate rightly emotion when the relevant event has passed a long time, but it have not yet helped us to represent the nature of continuous intensity of emotions. Additionally, this paper has formalized only the effect of trust/distrust on emotions but not yet the effect of emotions on trust/distrust. These current limitations are also the potential perspective for our future research.

Acknowledgements: This work has been supported by the Agence Nationale de la Recherche (ANR), contract No. ANR-08-CORD-005-1, and by a doctoral scholarship awarded by the University of Toulouse, contract No. 26977-2007.

Bibliography

[AACS08] P. Aghion, Y. Algan, P. Cahuc, A. Shleifer. Regulation and Distrust. *SUS.DIV-CEPR-PSEConference of Models of Cultural Dynamics and Diversity*, 2008.

- [AHL09] C. Adam, A. Herzig, D. Longin. A logical formalization of the OCC theory of emotions. *Synthese* 168(2):201–248, 2009.
- [Bry07] H. J. Bryce. Formalizing Civic Engagement: NGOs and the Concepts of Trust, Structure, and Order in the Public Policy Process. *Workshop on Building Trust Through Civic Engagement and for the International Political Science Association, Section on Governance, conference on Government Crisis in Comparative Perspective, Seoul, Korea*, 2007.
- [CES86] E. M. Clarke, E. A. Emerson, A. P. Sistla. Automatic verification of finite-state concurrent systems using temporal logic specifications. *ACM Transactions on Programming Languages and Systems* 8(2):244–263, 1986.
- [CF01] C. Castelfranchi, R. Falcone. Social Trust: A Cognitive Approach. In Castelfranchi and Tan (eds.), *Trust and Deception in Virtual Societies*. Pp. 55–90. Kluwer Academic Publishers, Dordrecht, 2001.
- [CFL08] C. Castelfranchi, R. Falcone, E. Lorini. A non-reductionist Approach to Trust. In Goldbeck (ed.), *Computing with Social Trust*. Pp. 45–72. Springer, Berlin, 2008.
- [CL90] P. R. Cohen, H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence* 42:213–261, 1990.
- [Dre95] F. Dretske. *Naturalizing the mind*. MIT Press, Cambridge, 1995.
- [FL79] M. Fischer, R. Ladner. Propositional dynamic logic of regular programs. *Journal of Computer and System Sciences* 18(2):194–211, 1979.
- [Fri86] N. H. Frijda. *The Emotions: Studies in Emotion & Social Interaction*. Edition de la Maison des Sciences de l’Homme. Cambridge University Press, Paris, 1986.
- [GG06] P. Gochet, P. Gribomont. Epistemic Logic. In Gabbay and Woods (eds.), *Twentieth Century Modalities*. Handbook of the History of Logic 7, pp. 99–195. Elsevier, amsterdam edition, 2006.
- [GHL06] B. Gaudou, A. Herzig, D. Longin. A Logical Framework for Grounding-based Dialogue Analysis. In Hoek et al. (eds.), *International Workshop on Logic and Communication in Multi-Agent Systems (LCMAS), Edinburgh, Scotland, UK, 01/08/2005*. Electronic Notes in Theoretical Computer Science (ENTCS) 157(4), pp. 117–137. Elsevier, <http://www.elsevier.com/>, 2006.
- [Hin62] J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca, 1962.
- [HKT00] D. Harel, D. Kozen, J. Tiuryn. *Dynamic Logic*. MIT Press, 2000.
- [HL04] A. Herzig, D. Longin. C&L intention revisited. In *Proceedings of Int. Conf. of knowledge representation and reasoning KR’04*. Pp. 527–535. Morgan Kaufmann, 2004.

- [HLH⁺08] A. Herzig, E. Lorini, J. F. Hübner, J. Ben-Naim, O. Boissier, C. Castelfranchi, R. Demolombe, D. Longin, L. Perrussel, L. Vercouter. Prolegomena for a logic of trust and reputation. In *Proceedings of 3rd International Workshop on Normative Multiagent Systems (NorMAS)*. Luxembourg, July 2008.
- [Lah01] B. Lahno. On the Emotional Character of Trust. *Journal of Ethical Theory and Moral Practice* 4:171–189, 2001.
- [Laz91] R. S. Lazarus. *Emotion & Adaptation*. Oxford University Press, 1991.
- [LW00] R. Lewicki, C. Wiethoff. Trust, Trust Development, and Trust Repair. In Deutsch and Coleman (eds.), *The Handbook of Conflict Resolution: Theory and Practice*. Pp. 86–107. Jossey-Bass, San Francisco, CA, 2000.
- [OCC88] A. Ortony, G. L. Clore, A. Collins. *The Cognitive Structure of Emotions*. The Cambridge University Press, 1988.
- [Pri57] A. N. Prior. *Time and Modality*. Clarendon Press, Oxford, 1957.
- [Rei09] R. Reisenzein. Emotions as metarepresentational states of mind: Naturalizing the belief-desire theory of emotion. *Cognitive Systems Research* 10(1):6–20, 2009.
- [Sch01] K. R. Scherer. *Appraisal Processes in Emotion : Theory, Methods, Research*. Chapter Appraisal Considered as a Process of Multilevel Sequential Checking, pp. 92–120. Oxford University Press, New York, 2001.
- [SDM07a] B. R. Steunebrink, M. Dastani, J.-J. C. Meyer. A Logic of Emotions for Intelligent Agents. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada*. Pp. 142–147. AAAI Press, 2007.
- [SDM07b] B. R. Steunebrink, M. Dastani, J.-J. C. Meyer. Towards a Quantitative Model of Emotions for Intelligent Agents. In Reichardt and Levi (eds.), *Proceedings of the 2nd Workshop on Emotion and Computing - Current Research and Future Impact*. Osnabrück, Germany, 2007.
- [dS01] R. de Sousa. *The Rationality of Emotion*. MIT Press, 6 edition, 2001.
- [SSJ01] K. R. Scherer, A. Schorr, T. Johnstone. *Appraisal Processes in Emotion: Theory, methode, Research*. Series in Affective Science. Oxford university Press, 2001.
- [WK95] D. N. Walton, E. C. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New-York Press, NY, 1995.