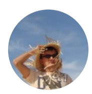
Relation of Trust and Social Emotions: A Logical Approach

Conference Paper · January 2009

DOI: 10.1109/WI-IAT.2009.166 · Source: DBLP

3 authors, including:

Dominique Longin
Institut de Recherche en Informatique de To…



Manh Hung Nguyen
Posts and Telecommunications Institute of T…

Relation of Trust and Social Emotions: A Logical Approach

Jean-François Bonnefon
Université de Toulouse
CNRS, CLLE, Toulouse, France
bonnefon@univ-tlse2.fr

Dominique Longin
Université de Toulouse
CNRS, IRIT, Toulouse, France
Dominique.Longin@irit.fr

Manh-Hung Nguyen
Université de Toulouse
UPS, IRIT, Toulouse, France
Manh-Hung.Nguyen@irit.fr

***Abstract*—Trust and social emotions such as gratitude and anger have natural relations and they both play a key role in research of interaction systems in the context of ambient intelligence and affective computing nowadays. This paper presents a logical approach to formalize both the relations between trust and anger for one hand, and between distrust and gratitude for another hand. Our formal framework is a multimodal logic that combines a logic of belief and choice, a logic of linear time, and a logic of norms. We also provide the behavioral validation for these relations.**

***Keywords*-Modal logic, cognitive structure of emotions, trust.**

I. 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 logical approach to represent formal relations between trust and cognitive structure of social emotions such as gratitude and anger. We refer to gratitude and anger as social emotions, because they depend on social standards and norms in addition to personal desires [1]. Although gratitude and anger have some natural relations with trust as well as distrust, and although trust [2] and emotions [3] have been separately formalized, there is not yet a formal representation for the relations between them. The purpose of our work is to represent within a logical framework the relations between social emotions and (dis)trust. We focus in particular on anger as consequence of the betrayal of trust; and on gratitude as a consequence of unwarranted distrust. The logic we offer is a combination of: a logic of belief and choice [4], [5], a logic of time [6]. A part of Dynamic Logic [7], [8] is built from logic of time.

The paper is organized as follow: Section II introduces the logical framework. Section III formalizes the concept of anger and gratitude. Section IV formalizes the concept of trust and distrust. Section V shows the relations between anger and trust, distrust and gratitude, and provides behavioral validation for these relations.

II. Logical Framework

Syntax. The syntactic primitives of our logic are as follows: a nonempty finite set of agents $AGT = \{i_1, i_2, \ldots, i_n\}$, a nonempty finite set of atomic events $EVT = \{e_1, e_2, \ldots, e_p\}$, and a nonempty set of atomic propositions $ATM = \{p_1, p_2, \ldots\}$. The variables i, j, $k\ldots$ 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, $\beta\ldots$ denote such actions. The language of our logic is defined by the following BNF :

$$\varphi ::= p \mid \bot \mid \textit{do-}i{:}\alpha \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathtt{X}\varphi \mid \mathtt{X}^{-1}\varphi \mid \mathtt{G}\varphi \mid \mathtt{Bel}_i\,\varphi \mid \mathtt{Choice}_i\,\varphi \mid \mathtt{Idl}_i\,\varphi$$

where p ranges over ATM, $i{:}\alpha$ ranges over $AGT \times EVT$ and $\textit{do-}i{:}\alpha$ ranges over ATM. The classical boolean connectives $\wedge$ (conjunction), $\rightarrow$ (material implication) and $\top$ (tautology) are defined from $\neg$ (negation), $\vee$ (disjunction) and $\bot$ (contradiction).

$\textit{do-}i{:}\alpha$ reads "agent i is just about to perform the action α"; $\mathtt{X}\varphi$ reads "φ will be true next time"; $\mathtt{X}^{-1}\varphi$ reads "φ was true at the previous time"; $\mathtt{G}\varphi$ reads "henceforth, φ is true"; $\mathtt{Bel}_i\,\varphi$ reads "agent i believes that φ is true"; $\mathtt{Choice}_i\,\varphi$ reads "agent i prefers that φ be true"; $\mathtt{Idl}_i\,\varphi$ reads "It is ideal for agent i to bring about φ". We define the following abbreviations:

$$\textit{done-}i{:}\alpha \overset{def}{=} \mathtt{X}^{-1}\textit{do-}i{:}\alpha \qquad (\text{Def}_{\textit{done-}i:\alpha})$$

$$\mathtt{After}_{i:\alpha}\varphi \overset{def}{=} \textit{do-}i{:}\alpha \rightarrow \mathtt{X}\varphi \qquad (\text{Def}_{\mathtt{After}_{i:\alpha}})$$

$$\mathtt{Done}_{i:\alpha}\varphi \overset{def}{=} \textit{done-}i{:}\alpha \wedge \mathtt{X}^{-1}\varphi \qquad (\text{Def}_{\mathtt{Done}_{i:\alpha}})$$

$$\mathtt{F}\varphi \overset{def}{=} \neg\mathtt{G}\neg\varphi \qquad (\text{Def}_{\mathtt{F}})$$

$$\mathtt{Goal}_i\,\varphi \overset{def}{=} \mathtt{Choice}_i\,\mathtt{F}\mathtt{Bel}_i\,\varphi \qquad (\text{Def}_{\mathtt{Goal}_i})$$

$$\mathtt{Int}_i\,(i{:}\alpha) \overset{def}{=} \mathtt{Choice}_i\,\mathtt{F}\textit{do-}i{:}\alpha \qquad (\text{Def}_{\mathtt{Int}_i})$$

$$\mathtt{Poss}_i\varphi \overset{def}{=} \neg\mathtt{Bel}_i\,\neg\varphi \qquad (\text{Def}_{\mathtt{Poss}_i})$$

$\textit{done-}i{:}\alpha$ reads "agent i has done action α"; $\mathtt{After}_{i:\alpha}\varphi$ reads "φ is true after any execution of α by i", and $\mathtt{After}_{i:\alpha}\bot$ reads "α cannot be performed by i"; $\mathtt{Done}_{i:\alpha}\varphi$ reads "agent i has done action α and φ was true at previous time"; $\mathtt{F}\varphi$ reads "φ will be true in some future instants"; $\mathtt{Goal}_i\,\varphi$ reads "agent i has the goal (chosen preference) that φ be true"; $\mathtt{Int}_i\,(i{:}\alpha)$ reads "agent i intends to do α"; $\mathtt{Poss}_i\varphi$ reads "agent i believes that φ is possible true".

Sometimes, we note $\neg done\text{-}i{:}\alpha$ as $done\text{-}i{:}{\sim}\alpha$ and $\neg do\text{-}i{:}\alpha$ as $do\text{-}i{:}{\sim}\alpha$. Thus, $\texttt{After}_{i:\sim\alpha}\varphi$ reads "if agent i does not performs α, next φ will be true". Similarly, we can write $\texttt{Done}_{i:\sim\alpha}\varphi$ and $\texttt{Int}_i\,(i{:}{\sim}\alpha)$.

Semantics. We use a semantics based on linear time described by a history of time points. (This semantics is very closed to CTL^* [9].) A **frame** $\mathcal{F}$ is a 4-tuple $\langle H, \mathscr{B}, \mathscr{C}, \mathscr{I}\rangle$:

- H is a set of histories that are represented as sequences of time points, where each time point is identified by an integer $z \in \mathbb{Z}$;
- $\mathscr{B}_i(h, z)$ denotes the set of histories believed as being possible by the agent i in the history h at the time z;
- $\mathscr{C}_i(h, z)$ denotes the set of histories chosen by the agent i in the history h at the time z;
- $\mathscr{I}_i(h, z)$ denotes the set of ideal histories for the agent i in the history h at the time z.

The **semantical constraints** are as follows: all the accessibility relations $\mathscr{B}_i$ are serial, transitive and euclidian; all the accessibility relations $\mathscr{C}_i$ are serial; all the accessibility relations $\mathscr{I}_i$ are serial; for every $z \in \mathbb{Z}$: if $h' \in \mathscr{B}_i(h, z)$ then $\mathscr{C}_i(h, z) = \mathscr{C}_i(h', z)$ and $\mathscr{I}_i(h, z) = \mathscr{I}_i(h', z)$.

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 defined as follows: $\mathcal{M}, h, z \models p$ iff $(h, z) \in \mathcal{V}(p)$; $\mathcal{M}, h, z \models \texttt{X}\varphi$ iff $\mathcal{M}, h, z+1 \models \varphi$; $\mathcal{M}, h, z \models \texttt{X}^{-1}\varphi$ iff $\mathcal{M}, h, z-1 \models \varphi$; $\mathcal{M}, h, z \models \texttt{G}\varphi$ iff $\mathcal{M}, h, z' \models \varphi$ for every $z' \geq z$; for every $z \in \mathbb{Z}$, $\mathcal{M}, h, z \models \texttt{Bel}_i\,\varphi$ iff $\mathcal{M}, h', z \models \varphi$ for every $h' \in \mathscr{B}_i(h, z)$; for every $z \in \mathbb{Z}$, $\mathcal{M}, h, z \models \texttt{Choice}_i\,\varphi$ iff $\mathcal{M}, h', z \models \varphi$ for every $h' \in \mathscr{C}_i(h, z)$; for every $z \in \mathbb{Z}$, $\mathcal{M}, h, z \models \texttt{Idl}_i\,\varphi$ iff $\mathcal{M}, h', z \models \varphi$ for every $h' \in \mathscr{I}_i(h, z)$. Other truth conditions are defined as usual.

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

$$do\text{-}i{:}\alpha \leftrightarrow \texttt{X}done\text{-}i{:}\alpha \quad (1) \qquad \varphi \leftrightarrow \texttt{X}^{-1}\texttt{X}\varphi \quad (4)$$

$$\texttt{X}\varphi \leftrightarrow \neg\texttt{X}\neg\varphi \quad (2) \qquad \texttt{G}\varphi \leftrightarrow \varphi \wedge \texttt{XG}\varphi \quad (5)$$

$$\varphi \leftrightarrow \texttt{XX}^{-1}\varphi \quad (3) \qquad \texttt{G}(\varphi \rightarrow \texttt{X}\varphi) \rightarrow (\varphi \rightarrow \texttt{G}\varphi) \quad (6)$$

$\texttt{Bel}_i$, $\texttt{Choice}_i$ and $\texttt{Idl}_i$ operators are defined in a normal modal logic plus ($D_\Box$), ($4_\Box$) and ($5_\Box$) axioms. Thus, if $\Box$ represents a $\texttt{Bel}_i$ or $\texttt{Choice}_i$ or $\texttt{Idl}_i$ operator:

$$\frac{\varphi}{\Box\varphi} \quad (RN_\Box)$$

$$\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi) \quad (K_\Box)$$

$$\Box\varphi \rightarrow \neg\Box\neg\varphi \quad (D_\Box)$$

$$\Box\varphi \leftrightarrow \texttt{Bel}_i\,\Box\varphi \quad (4_\Box)$$

$$\neg\Box\varphi \leftrightarrow \texttt{Bel}_i\,\neg\Box\varphi \quad (5_\Box)$$

($RN_\Box$) means that all theorems are believed (chosen, ideal) by every agent i; ($K_\Box$) means that beliefs (choices, ideals) are closed under material implication for every agent i; ($D_\Box$) means that beliefs (choices, ideals) of every agent i are rational: they cannot be contradictory. ($4_\Box$) and ($5_\Box$) mean that agent i is conscious of its beliefs (choices, ideals) and of its disbeliefs (no-choices, not-ideals respectively).

Finally, we note for convenience $\texttt{Idl}_i\,do\text{-}j{:}\alpha$ as $\texttt{Idl}_i\,(j{:}\alpha)$.

III. Formalization of Cognitive Structure of Social Emotions

In this section, we present the formalization of social emotions, such as gratitude and anger, based on their cognitive structure as proposed by Ortony *et al.* [1].

Gratitude. According to Ortony *et al.*, *Gratitude*'s cognitive structure has two factors: (i) agent i believes that agent j has done a praiseworthy (ideal) action α and that the result φ of such an action is desired by agent i: $\texttt{Bel}_i\,\texttt{Done}_{j:\alpha}(\texttt{Idl}_j\,(j{:}\alpha) \wedge \texttt{Bel}_i\,\texttt{After}_{j:\alpha}\varphi \wedge \texttt{Goal}_i\,\varphi)$; and (ii) agent i now experiences that in fact φ is true: $\texttt{Bel}_i\,\varphi$. Thus, we accordingly formalize the concept of *Gratitude* as:

***Definition** 1 (Gratitude):*

$$\texttt{Gratitude}(i, j, j{:}\alpha, \varphi) \overset{def}{=} \texttt{Bel}_i\,\varphi \wedge$$
$$\texttt{Bel}_i\,\texttt{Done}_{j:\alpha}(\texttt{Goal}_i\,\varphi \wedge \texttt{Idl}_j\,(j{:}\alpha) \wedge \texttt{Bel}_i\,\texttt{After}_{j:\alpha}\varphi)$$

Anger. *Anger*'s cognitive structure also has two factors: (i) agent i believes that agent j has done a blameworthy action (it is ideal to not do such an action) α and agent i had no desire (formalized as desire of $\neg\varphi$) for the action's outcome (for i, it is possible to avoid φ by mot doing α): $\texttt{Done}_{j:\alpha}(\texttt{Idl}_j\,(j{:}\sim\alpha) \wedge \texttt{Goal}_i\,\neg\varphi \wedge \texttt{Poss}_i\texttt{After}_{j:\sim\alpha}\neg\varphi)$; and (ii) agent i now experiences the unexpected outcome of such an action: $\texttt{Bel}_i\,\varphi$. We accordingly formalize the concept of *Anger* as:

***Definition** 2 (Anger):*

$$\texttt{Anger}(i, j, j{:}\alpha, \varphi) \overset{def}{=} \texttt{Bel}_i\,\texttt{Done}_{j:\alpha}(\texttt{Goal}_i\,\neg\varphi \wedge$$
$$\texttt{Idl}_j\,(j{:}\sim\alpha) \wedge \texttt{Poss}_i\texttt{After}_{j:\sim\alpha}\neg\varphi) \wedge \texttt{Bel}_i\,\varphi$$

In this formula, if we replace $j{:}\alpha$ by $j{:}\sim\alpha$, it becomes $\texttt{Anger}(i, j, j{:}\sim\alpha, \varphi)$, reads "agent i is angry about agent j to not do α in order to avoid φ." In other terms, this formula corresponds to the concept of anger in inaction.

IV. Formalization of Trust

We now present the formalization of trust and distrust based on the cognitive definition of Castelfranchi and colleagues [10], [11].

Trust. Castelfranchi and Falcone's definition [10] of trust in action says that agent i trusts agent j to ensure φ by performing action α if and only if agent i: desires to achieve φ ($\texttt{Goal}_i\,\varphi$); believes that φ can be achieved by doing action α ($\texttt{Bel}_i\,\texttt{After}_{j:\alpha}\varphi$); expects that agent j is capable to perform action α ($\neg\texttt{Bel}_i\,\texttt{After}_{j:\alpha}\bot$); and believes that agent j has the intention to do action α ($\texttt{Bel}_i\,\texttt{Int}_j\,(j{:}\alpha)$). We accordingly formalize this concept as:

Definition 3 (Trust):

$$\texttt{Trust}(i, j, j{:}\alpha, \varphi) \stackrel{def}{=} \texttt{Goal}_i\, \varphi \wedge \texttt{Bel}_i\, \texttt{After}_{j:\alpha}\varphi \wedge \neg\texttt{Bel}_i\, \texttt{After}_{j:\alpha}\bot \wedge \texttt{Bel}_i\, \texttt{Int}_j\, (j{:}\alpha)$$

Note that in case of trust in inaction, we can simply replace $j{:}\alpha$ in this formula by $j{:}\sim\alpha$, we accordingly have the concept of trust in inaction: $\texttt{Trust}(i, j, j{:}\sim\alpha, \varphi)$. This formula can be translated as: agent i trusts agent j to avoid (not to do) action α to maintain/achieve φ.

Distrust. We also adopt the definition of distrust given by Castelfranchi et al. [11] 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 ($\texttt{Goal}_i\, \varphi$), and agent i believes that at least one of these conditions is fullfiled: (i) agent j is not in the capacity to do action α to bring about φ: $\texttt{Bel}_i\, \neg\texttt{After}_{j:\alpha}\varphi$; or (ii) agent j is able to do α but he has no intention to do α: $\texttt{Bel}_i\, \texttt{After}_{j:\alpha}\varphi \wedge \texttt{Bel}_i\, \texttt{Int}_j\, (j{:}\sim\alpha)$. We accordingly formalize this concept as:

Definition 4 (Distrust):

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

V. Trust-Related Social Emotions

A. Formal Relations

Trust – Anger. Betrayal is one of the most important features of trust: Trust has a posibility of being betrayed [10]. After being betrayed, we are likely to feel anger. We accordingly formalize this relation as:

Proposition 5 (Betrayal of Trust implies Anger):

$$\texttt{Bel}_i\, \texttt{Done}_{j:\alpha}(\texttt{Trust}(i, j, j{:}\sim\alpha, \neg\varphi) \wedge \texttt{Idl}_j\, (j{:}\sim\alpha)) \wedge \texttt{Bel}_i\, \varphi \rightarrow \texttt{Anger}(i, j, j{:}\alpha, \varphi)$$

In other terms, we feel angry when we trust someone to avoid an action that is ideal to avoid (in order to achieve/maintain some outcomes), but discover that this person performed the action, and that we thus experience some unexpected outcome.

Imagine that you are driving in a one-way street, and get hit by another car that came from the wrong direction, damaging your own car. You feel anger ($\texttt{Anger}(you, driver,$*driver: come from wrong direction*,*your car is damaged*$)$) because you trusted other drivers not to come from the wrong direction, in order to prevent your car from being damaged ($\texttt{Trust}(you, driver, driver:$ $\neg$*(come from wrong direction)*, $\neg$*(your car is damaged)*$)$); It is normally ideal to avoid driving in the wrong direction ($\texttt{Bel}_{you}\, \texttt{Idl}_{driver}\, \neg$*(come from wrong direction)*). But that driver did come from the wrong direction, and this action caused the unexpected outcome of damaging your car ($\texttt{Goal}_{you}\, \neg$*(your car is damaged)*).

Note that in this proposition, if we replace $j{:}\alpha$ by $j{:}\sim\alpha$, it becomes:

$$\texttt{Bel}_i\, \texttt{Done}_{j:\sim\alpha}(\texttt{Trust}(i, j, j{:}\alpha, \neg\varphi) \wedge \texttt{Idl}_j\, (j{:}\alpha)) \wedge \texttt{Bel}_i\, \varphi \rightarrow \texttt{Anger}(i, j, j{:}\sim\alpha, \varphi)$$

In other terms, this formula corresponds to the case of anger in inaction which comes from trust in action. Thus, based on the single idea of anger coming from the betrayal of trust, Proposition (Prop.) 5 covers two cases: anger in action comes from the betrayal of trust in inaction; and anger in inaction comes from the betrayal of trust in action.

DisTrust – Gratitude. From definition 4 which says that distrust can base upon the non capacity or non intention, we can deduce for the case of non-intention-based distrust as:

Definition 6 (Non-intention-based Distrust):

$$\texttt{I-DisTrust}(i, j, j{:}\alpha, \varphi) \stackrel{def}{=} \texttt{Goal}_i\, \varphi \wedge \texttt{Bel}_i\, \texttt{After}_{j:\alpha}\varphi \wedge \texttt{Bel}_i\, \texttt{Int}_j\, (j{:}\sim\alpha)$$

We accordingly formalize the relation between non-intention-based distrust and gratitude as:

Proposition 7: (Unconfirmation of non-intention-based DisTrust implies Gratitude)

$$\texttt{Bel}_i\, \texttt{Done}_{j:\alpha}(\texttt{I-DisTrust}(i, j, j{:}\alpha, \varphi) \wedge \texttt{Idl}_j\, (j{:}\alpha)) \wedge \texttt{Bel}_i\, \varphi \rightarrow \texttt{Gratitude}(i, j, j{:}\alpha, \varphi)$$

In other terms, we feel grateful about an action when we distrusted someone to do an ideal (praiseworthy) action that would deliver desirable results (because we believed that person had no intention to do the action); but we then discover that this person actually did perform the action, and that we now experience our expected outcome.

B. 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 two propositions in this article, related to the emotions that follow trust when it is betrayed (Prop. 5); and the emotions that follow distrust, when it is unconfirmed (Prop. 7).

A total of 100 participants took part in an online survey. 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). Following Castelfranchi et al. [10], [11] who argue that trust is the conjunction of the intention and the capacity of trustee (see Section IV), we operationalize *Trust* as the conjunction of *Intention* and *Capacity*, and *Distrust* as the three remaining cases. Two other variables, the trustor's goal and consequence relation between α and its outcome φ, are fixed because they are the same in both trust and distrust's definition. As an example,

Table I
MEAN AND STANDARD DEVIATIONS OF AFFECTIVE RATINGS, AS A FUNCTION OF TRUST AND OUTCOME

	Anger		Gratitude	
	Trust	Distrust	Trust	Distrust
Success	1.1 (0.3)	1.2 (0.8)	3.7 (1.7)	4.3 (1.7)
Failure	3.9 (1.8)	2.7 (1.5)	1.2 (0.6)	1.4 (0.9)

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?

After reading each story, participants rated the extent to which the main character would feel each of 7 emotions including anger and gratitude. Ratings used a 6-point scale anchored at *Not at all* and *Totally*.

Descriptive statistics are displayed in Table I. 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.

Anger. The analysis detected main effects of Trust, $F(1, 98) = 22.4$, $p < .001$, $\eta 2_p = .18$; and Outcome, $F(1, 98) = 245$, $p < .001$, $\eta 2_p = .71$. However, these main effects were qualified by an interaction effect Trust $\times$ Outcome, $F(1, 98) = 37.5$, $p < .001$, $\eta 2_p = .28$. This interaction reflects our expectation (Prop. 5) that anger is especially high when trust is betrayed.

Gratitude. The analysis of variance detected a large effect of Outcome, $F(1, 98) = 272$, $p < .001$, accounting for most of the observed variance, $\eta 2_p = .73$. In other terms, Gratitude is largely predicted by Outcome alone. The analysis, however, also detected a comparatively small interaction effect involving Outcome $\times$ Trust, $F(1, 98) = 3.2$, $p < .01$, $\eta 2_p = .03$, reflecting the fact that apositive outcome has an even greater effect in case of distrust.

To get a more precise understanding of this interaction, we decomposed distrust in the factors Intention and Capacity, and analyzed their effects on gratitude after a positive outcome. This analysis detected a large effect of Intention, $F(1, 79) = 20.5$, $p < .001$, but no significant effect of Capacity, $F(1, 79) = 0.3$, $p < .59$. In other terms, and in consistent fashion with Prop. 7, gratitude is greater when distrust was based on Intention, rather than on Capacity.

VI. CONCLUSION

In this paper, we introduced a logical framework to represent trust, distrust and social emotions such as anger and gratitude. This logic enabled us to formalize the relations between trust and social emotions: anger comes from the betrayal of trust, and gratitude comes from unwarranted distrust. These formal relations were validated by a behavioral investigation 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 social emotions.

ACKNOWLEDGMENT

This work has been supported by the Agence Nationale de la Recherche, contract No. ANR-08-CORD-009.

REFERENCES

[1] A. Ortony, G. L. Clore, and A. Collins, *The Congnitive Structure of Emotions*. The Cambridge University Press, 1988.

[2] A. Herzig, E. Lorini, J. F. Hübner, J. Ben-Naim, O. Boissier, C. Castelfranchi, R. Demolombe, D. Longin, L. Perrussel, and L. Vercouter, "Prolegomena for a logic of trust and reputation," in *Proceedings of 3rd International Workshop on Normative Multiagent Systems (NorMAS)*, Luxembourg, July 2008.

[3] C. Adam, A. Herzig, and D. Longin, "A logical formalization of the OCC theory of emotions," *Synthese*, vol. 168, no. 2, pp. 201–248, 2009.

[4] A. Herzig and D. Longin, "C&L intention revisited," in *Proceedings of Int. Conf. of knowledge representation and reasoning KR'04*. Morgan Kaufmann, 2004, pp. 527–535.

[5] P. R. Cohen and H. J. Levesque, "Intention is choice with commitment," *Artificial Intelligence Journal*, vol. 42, no. 2–3, pp. 213–261, 1990.

[6] A. N. Prior, *Time and Modality*. Oxford: Clarendon Press, 1957.

[7] M. Fischer and R. Ladner, "Propositional dynamic logic of regular programs," *Journal of Computer and System Sciences*, vol. 18, no. 2, pp. 194–211, 1979.

[8] D. Harel, D. Kozen, and J. Tiuryn, *Dynamic Logic*. MIT Press, 2000.

[9] E. M. Clarke, E. A. Emerson, and A. P. Sistla, "Automatic verification of finite-state concurrent systems using temporal logic specifications," *ACM Transactions on Programming Languages and Systems*, vol. 8, no. 2, pp. 244–263, 1986.

[10] R. Falcone and C. Castelfranchi, "Social trust: a cognitive approach," in *Trust and Deception in Virtual Societies*, C. Castelfranchi and Y.-H. Tan, Eds. Kluwer Academic Publisher, 2001, pp. 55–90.

[11] C. Castelfranchi, R. Falcone, and E. Lorini, "A non-reductionist approach to trust," in *Computing with Social Trust*, J. Goldbeck, Ed. Berlin: Springer, 2008, pp. 45–72.