University of Toulouse

University of Toulouse II Le Mirail

PhD in computer sciences / artificial intelligence

A Logical Framework for Trust-Related Emotions:

Formal and Behavioral Results

by

Manh Hung NGUYEN

Co-supervisors:

Jean-François BONNEFON

Dominique LONGIN

Toulouse, September 2010

Abstract

Recently, emotions play a more and more important role in intelligent behavior researching in Artificial Intelligent area, specially in intelligent interaction systems. Purposes of these researches are to construct natural interactions between human and machine as those among humans by assigning computers human-like abilities of observation, interpretation (understanding, representation, reasoning) and generation of affect features, including user's emotions. This thesis proposes a logical framework which enables computer to represent emotions as well as to reason about emotions based upon some causal attitude which has a tight and natural relation with emotions: trust and distrust mental states. In other terms, from our proposed logical framework, we formalize the concept of some emotions as well as trust and distrust which are composed of their cognitive factors as argumentation of cognitive psychologists. From these formal concepts, we show that there are some relations between trust/distrust and emotions, particularly the effect of trust and distrust on emotions: a success following trust brings about satisfaction and/or gratitude; a failure following trust brings about disappointment and/or anger; a success following distrust brings about relief and/or gratitude; and a failure following distrust brings about fear-confirmed. These relations, after being formalized at the formal level, will also be validated by collecting empirical data, at the behavioral level.

Remerciements

Sachant que les mots ne peuvent exprimer la gratitude de mon profond respect pour ceux qui m'ont aidé à terminer la thèse, je tiens à les remercier du fond du coeur.

Mes premiers remerciements vont à Andrea Herzig et Dominique Longin qui m'ont donné la chance d'effectuer cette thèse. Je remercie également Andrea Herzig de me permettre de suivre son cours de logique pour M2R à l'UPS, de lire la thèse et puis, de me donner les utiles corrections et commentaires.

Je remercie également mes encadrants, Jean-François Bonnefon et Dominique Longin, pour m'avoir encadré et supporté pendant trois ans, pour leur grande disponibilité, leurs encouragements et conseils qu'ils m'ont prodigués.

Je remercie mes rapporteurs, Sylvie Pesty et Christiano Castelfranchi, pour avoir accepté de relire mon manuscrit, de me poser les questions et d'assister au jury de ma soutenance (je souhaite).

Merci à Philippe Balbiani pour les guides sur la sémantique de la logique et les instructions pour prouver la décidabilité de la logique.

Merci à Emiliano Lorini pour les discussions intéressantes, pour m'aider à inviter Christiano Castelfranchi comme un rapporteur de ma thèse, pour avoir partagé le bureau avec moi pendant deux premières années de thèse.

Merci à mon ancien enseignant Dinh Que Tran pour m'avoir aidé à relire mon manuscrit, pour les corrections, les commentaires et les discussion à distance. Je le remercie également pour m'avoir supporté et encouragé tout au long de la thèse.

Merci à Sylvie Leblois pour m'aider à corriger les questionnaires en français pour toutes les expériences que j'ai menées.

Merci à Fédéric Moisan pour m'avoir aidé à relire mon manuscrit, pour les corrections et les commentaires.

Merci aux membres du séminaire hebdomadaire en cognition, Jean-François Bonnefon, Sylvie Leblois, Stéphanie Pighin, Marco Heimann, Bastien Trémolière, Virgine Demeure, Charles Pompanon et Eric Billaut, qui m'a aidé à améliorer ma qualité de présentation.

Merci à Kourjieh Mounira et à Madeleine El Zaher pour avoir partagé le bureau avec moi pendant la troisième année de thèse. Merci à Kourjieh Mounira pour m'avoir partagé son expérience des procédures de dépose des dossiers de soutenance.

Merci aux membres de l'équipe LILaC, du laboratoire IRIT (à l'Université Toulouse III Paul Sabatier) et CLLE-LTC (à l'Université Toulouse II Le Mirail) qui m'ont accueilli et auprès de qui j'ai pu expérimenter la vie de laboratoire.

Merci à ma famille pour m'avoir permis d'arriver jusqu'ici. Merci de votre confiance et de votre soutien.

Enfin, je remercie l'allocation de recherche distribué par l'Université de Toulouse dont le numéro de contrat est de No. 26977-2007 ayant permis de financer pendant trois ans ma thèse. Je remercie également le projet ANR dont le numéro de contrat est de No. ANR-08-CORD-005-1 pour m'avoir financé pendant les conférences/workshops internationaux auxquels j'ai participé pendant ma thèse.

Contents

List of Symbols and Abbreviations

Abbreviation	Description	Definition
AmI	Ambient Intelligence	page 28
AI	Artificial Intelligence	page 1
ASL	Authorization Specification Language	page 56
BDI	Belief-Desire-Intention	page 36
BDTE	Belief-Desire Theory of Emotion	page 21
FIPA	Foundation for Intelligent Physical Agents	page 56
HMM	Hidden Markov Model	page 34
NPC	Non Player Characters	page 29
TTS	Text to Speech	page 32

List of Figures

List of Tables

Introduction

Recently, emotions play a more and more important role in intelligent behavior researching in Artificial Intelligence (AI). There are some motivations for reason why emotions become an active research area in computer science. First, having emotional abilities enable agents more believable. The AI researchers try to create the illusion of life by endowing agents with human abilities that they consider essential in intelligence (like reasoning, problem solving, learning, etc. cf. Bates [15]). Gratch and Marsella [96] agree on the crucial role emotions play in human intelligence and on the need to integrate them into virtual agents. However they notice that computer scientists are mainly interested in making their agents more believable by endowing them with emotional expressions. Second, emotions have an impact on cognition. Many psychologists proposed models of the impact of emotions on cognitions. For instance, Forgas [79] proposes the Affect Infusion Model (AIM), a framework allowing to explain how affectively loaded information can modify reasoning. Others researches attest the impact of emotions on abilities characteristic of human intelligence (planning, memory, learning, etc. cf. Gratch and Marsella [96]). Third, that is because of the birth of affective computing (Picard [195]) whose objective is to construct a transparent and adaptive interface between human and machine. In order to do this, the computer must focus not only on user's input, but also on the user's emotional states. This gives computers the human-like abilities of observation, interpretation and generation of affect features, including user's emotions. The ability of interpretation seem to be at the core of emotional processing because it requires technologies which enable computers to understand, represent and reason about emotions.

In order to construct an emotional model that would fulfill this potential requirement, computer scientists capitalize on psychological theories of emotion. There are many theories of emotions in psychology (de Sousa [235]), such as feelings theories which argue that emotions are feelings caused by changes in physiological conditions relating to the autonomic and motor functions (James-Lange [119], Cannon-Bard [36, 37], Schacter-Singer [219]); appraisal theories which consider emotion as a subjective evaluation of the environment in relation to the agent's goals (Arnold [9], Frijda [81], Lazarus [135], Roseman and Smith [208]), and cognitive theories which say that emotions are characterized by their associated cognitions and therefore, can be described as a set of beliefs, desires and feelings (Ortony et al. [186], Oakley [178], Reisenzein [205]). Most emotional applications for computers are built on the appraisal and cognitive theories of emotions, which assume that emotions are closely tied to changes in beliefs and desires such as the cognitive structure of emotion of Ortony et al. [186], the cognitive pattern of emotion of Lazarus [135] and the belief-desire theory of emotion (BDTE) of Reisenzein [205].

The appraisal and cognitive theories of emotions seem to be suitable to models constructed for interaction systems because: (i) their position on the awareness of emotions enables agents to recognize their emotion as well as that of their partners, and (ii) these theories assume that each emotion is characterized by its proper cognitive structure, this feature enables the modeler to construct emotions by considering primitive cognitive operators, such as belief and desire. There are many models of emotions based on appraisal and cognitive theories, from computational models such as that of Gershenson [87] which handles the degree of contradiction that emotions might have, the model of Hu et al. [111] based on the Belief-Desire-Intention (BDI) model [86], or Steunebrink et al.'s [238, 239] attempt to combine the qualitative and quantitative aspects of emotions. Let us also mention the formal models of Sanders [218], Meyer [164] which is also based on the BDI model, the framework of dynamic logic and (an extension of) the KARO framework [249], the model of Ochs et al. [181, 182], and the model of Adam et al. [3] which is also based on BDI model.

This thesis proposes a logical framework which enables computers to represent emotions, based on their cognitive factors. We also attempt to give the computer the ability to reason about emotions, based on some attitude which has a tight and natural relation with emotions: trust and distrust mental states.

In order to formalize the concept of trust for rational intelligent agents, we capitalize on the definitions of trust/distrust which argues that trust/distrust have their proper cognitive factors, conscious and subjective. We work from the cognitive definition of Castelfranchi and colleagues [39, 41, 75] who argue that trust is a combination of various beliefs: competence belief, dependence belief, disposition belief (which is decomposed into willingness belief and persistence belief), self-confidence belief, and motivation belief. The idea that trust is a combination of beliefs about the ability, intention, and disposition of the trustee is the main foundation of many models of trust in computer applications: computational model such as Marsh [153], Schillo et al. [223], Abdul-Rahman and Hailes. [1], Esfandiari and Chandrasekharan [72], Lashkari et al. [134], Yu and Singh [262, 263, 264], Sen and Sajja [228], Carter et al. [38], Manchala [149], Nefti et al. [173] and Ramchurn et al. [202]; or formal model such as Jones and Firozabadi. [123], Rangan [204], the Epistemic Event Temporal Trust (E2T2) model of Grandison and Reichgelt [92] and the logical model of Herzig et al. [103, 104].

Research has pointed out tight conceptual connections between emotion and trust, at the theoretical level for Lahno [131], Barbalet [13], Bryce [32], Aghion et al. [6], Lazarus [135, 136], Lewicki and Wiethoff [139]; or at the behavioral level for Fehr et al. [76] among others. However, although there are some separated formalizations of the concept of trust (Herzig et al. [103, 104], Jones and Firozabadi [123], and Rangan [204]), and of the concept of emotions (Adam et al. [3], Meyer [164], Ochs et al. [181, 182], Sanders [218], Steunebrink et al. [238, 239]), 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 which is a combination of the logic of beliefs and choices of Herzig and Longin [105] (a refinement of Cohen and Levesque [50]), and the logic of time (introduced by Arthur Prior [201]). This common logic will enable us to lay bare formal relations between trust and emotion, particularly the effect of trust and distrust on emotions: a success following trust brings about satisfaction and/or gratitude; a failure following trust brings about disappointment and/or anger; a success following distrust brings about relief and/or gratitude; and a failure following distrust brings about fear-confirmed. These effect of trust and distrust on emotions, after being formalized at the formal level, will also be behaviorally validated by collecting some empirical data.

This thesis is composed of seven chapters which are organized as follows:

Chapter 1 introduces a selective overview of researches on emotion in psychology, including three main theories: theory of feeling, appraisal theory and cognitive theory.

Chapter 2 presents a general view of researches on emotions in computer science. Particularly, we examine the research on the treatment of emotion in affective computing.

Chapter 3 offers a mixed overview of trust research, in psychology and computer science. Particularly, we examine definitions of trust and related concepts. Then, we try to categorize research on trust in computer science into three groups: protocol-oriented models, computational models, and formal models.

Chapter 4 introduces our logical framework, as a common logic for emotions and trust/distrust; this common logic will enable us to lay bare formal relations between trust and emotion, particularly the effect of trust and distrust on emotions. The logic we offer is a combination of the logic of beliefs and choices, and the logic of time.

Chapter 5: Using the logic introduced in chapter 4, this chapter presents the formalization of the cognitive structure of emotion, including well-being emotions such as joy and distress; prospect-based emotions such as hope and fear; confirmation-based emotions such as satisfaction, disappointment, fear-confirmed and relief; and compound emotions such as gratitude and anger. This chapter also presents the formalization of the cognitive structure of trust and distrust in the same logic.

Chapter 6: From the formal representation of the cognitive structure of emotion, trust, and distrust introduced in chapter 5, this chapter represents the effect of trust and distrust on emotion, in the same logic introduced in chapter 4, in the form of propositions.

Chapter 7 offers validation by behavioral investigations, following the methods of experimental psychology. The success of these behavioral validations gives strong support to our approach, which is shown to capture lay users' intuitions about the effect of trust and distrust on emotions.

Chapter 1

Emotion in Psychology

This chapter introduces a selective overview of researches on emotion in psychology. Section 1.1 presents a brief history of researches on emotion; section 1.2 presents feeling theories of emotion; section 1.3 presents appraisal theories of emotion; section 1.4 presents theories of emotion in the cognitive point of view; section 1.5 presents relations between emotion and its related term: mood.

1.1 A brief history

Aristotle's analysis of emotions in the *Rhetoric* is considered as the earliest systematic discussion of human psychology [69]. It is valuable both for the light it throws on the emotions in general and for what it says about the role of the emotions in Greek political life.

After the Aristotle's period, the world had to await until the 19th century with Darwin's book *Expression of Emotion in Man and Animals* (1872). It is considered as the first study of the emotions using scientific methodology (cf. Elster [69]). Darwin's book examines how people and animals display varieties of emotions via their facial expressions on three general principles [54]: The principle of serviceable associated habits; the principle of antithesis; and the principle of actions due to the constitution of the nervous system.

Late 19th century is also the birth time of feeling theories. In 1884, James [119] introduced his theory. Then, Cannon [36, 37] in 1927 and 1929, Schacter-Singer [219] in 1962 respectively proposed their theory. These theories argue that emotion is a class of feelings: emotions are triggered by the perception of the physiological changes of body. While Darwin treated only external signals of emotions via facial expressions, James, Cannon and Schacter also took into account internal signals of emotions: feelings in the physiological change of body. This could be considered as a positive point of feeling theories at that time. However, problems of these theories, which are in ability to distinguish emotions, in rationality of emotions, and in respecting the fact that emotions are directed at intentional objects, encourage researches to review and understand emotions more deeply.

A different views of emotions is to regard to their functional interpretation. In 1949, Hebb [101] considered emotions as disruptions or disturbances because emotions are caused by events to which the organism can not adapt. However, such a view is obviously opposite to the view that emotions have adaptive functions (cf. Damasio [53]).

The distinction that emotions have disruption or adaptive functions is also the origin of the trend to evaluate what happens to the owner's goodwill, which is the main idea of appraisal theories. In 1960, Arnold [9] argued that emotions are a kind of attraction (something appraised as good) or aversion (something appraised as bad). In 1975, Taylor [243] argued that emotions are rational (about what caused them) which is then confirmed by de Sousa [56] in 2001. In 1983, Mandler [150] considered emotions as states of autonomic arousal. In 1986, Frijda [81] viewed emotions as outcomes of the process of assessing the world in term of individual's concerns. In 1991, Lazarus [135] argued that emotions arise from an adaption of the individual to the environment. In 2003, Whisner [257] also considers that emotions involve evaluations.

While appraisal theories do not explicitly mention conscious and propositional properties of emotions' content, cognitive theories claim that emotions are conscious and that they involve propositional attitudes. This enhancement enables to recognize and distinguish emotions. In 1988, Ortony et al. [186] proposed their theory of emotion which concentrates on cognitive elicitors, including events, agents and objects. Each emotion has an unique combination of these elicitors' values, this combination is called the cognitive structure of such an emotion. In 2009, Reisenzein [205] introduced a theory which assumes that emotions is a result of a process to sensory changes in the state of the belief-desire system.

The cognitive theories' ability, which enables to identify emotions, brings a new generation and tendency in emotional research: make up the computer abilities to understand, represent and reasoning about human emotions. In 1997, Picard [195] introduced a concept *affective computing*. She assumes that some computer systems may need to be able to recognize the user's emotions. Researches about emotions in computer application will be presented in chapter 2.

The following sections will give an overview on the main trends in the theorization of emotion, including theories of feeling, appraisal theories and cognitive theories.

1.2 Feeling theories

The main idea of these theories is that emotions are triggered by the perception of the physiological changes in the body. For instance, when we see a snake, we tend to seek safety in flight. Our flight is interpreted as the danger of the snake, therefore we feel fear (cf. James [119]). This kind of theory is considered as the most representative of common sense, differentiated from perceptions by their experienced quality (de Sousa [235]).

We present here three theories in this category: the theory of James-Lange [119], the theory of Cannon-Bard [36, 37] and the theory of Schacter-Singer [219]. We also discuss on some limits of theories in this category.

1.2.1 The theory of James-Lange

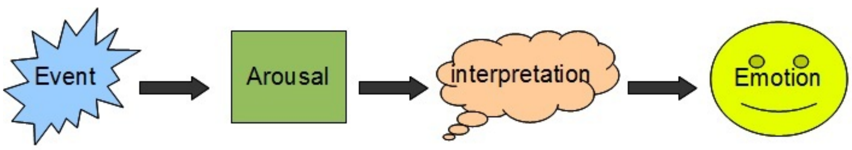


Figure 1.1: The sequence leading to emotion in the theory of James-Lange

James [119] proposed a theory (commonly known as the "James-Lange" theory of emotion, after James and Carl G. Lange) according to which an event first causes physiological arousal and physical response. It is not an emotion until the subject interprets the physical response as an emotion. For instance, we feel sorry because we cry, angry because we strike, or afraid because we shake. Basically what this is saying is that the physical aspects appear before the emotions are perceived. So until we physically respond to a situation we do not perceive or understand the emotions surrounding it (Figure 1.1).

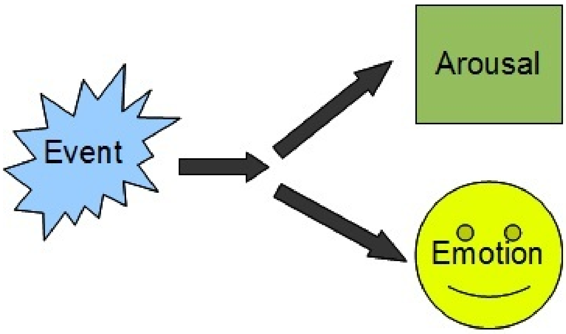


Figure 1.2: The sequence leading to emotion in the theory of Cannon-Bard

1.2.2 The theory of Cannon-Bard

The Cannon-Bard theory was originally developed by Walter Cannon [36, 37] and later expanded upon by physiologist Phillip Bard. The Cannon-Bard theory suggests the following chain of events: emotion-provoking stimuli are received by the senses and are then relayed simultaneously to the cerebral cortex, which provides the conscious mental experience of the emotion, and to the sympathetic nervous system, which produces the physiological state of arousal. Basically what this is saying, is that we experience both the physical aspect and the perception of the emotion at the same time, rather than one being caused by the other (Figure 1.2).

1.2.3 The theory of Schacter-Singer

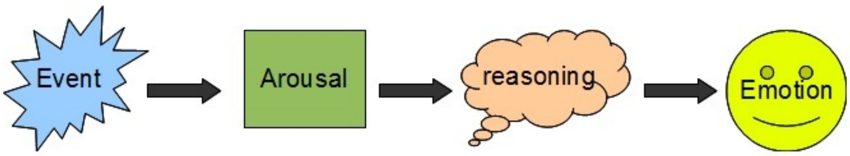


Figure 1.3: The sequence leading to emotion in the theory of Schacter-Singer

The theory of Schacter-Singer [219] is a two-stage theory. They state that for an emotion to occur there must first be physiological arousal, and second there must be an explanation for the arousal (Figure 1.3). The chain of events is similar to those of James-Lange's theory where interpretation phase is replaced by reasoning phase.

1.2.4 Limits of feeling theories

The first problem with this kind of theory is that it is unable to distinguish between different emotions (cf. Cannon [37], de Sousa [235]). These theories have not any factor or criteria which enable to differentiate emotions: there is only the feeling of an individual. Although James [119] argues what distinguishes emotions is the fact that each involves the perception of a unique set of bodily changes, the characteristic or feature of such a set is not easy to identify. For instance, Cannon [37] claimed that the reactions feature of fear and anger are identical, and so these reactions can not allow us to identify what emotions occur. The same conclusion is usually drawn from an experiment performed by Schacter and Singer [219] which independently manipulated physiological and situational determinants of emotional state. It was demonstrated that neither physiological nor psychological explanations alone could account for the experimental facts.

The second problem with feeling theories is that emotions are irrational. They suggest to treat emotions as brute facts: in spite biological or psychological features can explain emotions, they can not be rationalized (cf. de Sousa [235]). Emotions, however, are capable of being not only explained, but also justified - they are closely related to the reasons that give rise to them. For instance, Taylor [243] argues that if someone angers me, I can cite my antagonist's deprecatory tone; if someone makes me jealous, I can point to his poaching on my property.

The last problem with this kind of theory is that it assimilates emotions to sensations in spite of their difference in experienced quality. Therefore, feeling theories failed to take account of the fact that emotions are typically directed at intentional objects, which is only taken in appraisal and cognitive theories.

1.3 Appraisal theories

Appraisal theories say that emotions typically have formal objects which highlight another important feature of emotional experience which feeling theories neglect, and which other psychological theories attempt to accommodate: emotions involve evaluations. For instance, we feel fear about the snake because we think (evaluate) that the snake is dangerous (cf. Whisner [257]).

Magna Arnold [9] introduced the notion of appraisal into psychology, characterizing it as the process through which the significance of a situation for an individual is determined. Appraisal gives rise to attraction or aversion, and emotion is equated with this felt tendency toward anything intuitively appraised as good (beneficial), or away from anything intuitively appraised as bad (harmful).

Frijda [81] argues that emotions arise because events are appraised by people as favorable or harmful to their own interests. He takes an information-processing perspective: Emotions are viewed as outcomes of the process of assessing the world in terms of one's own concerns, which, in turn, modify action readiness.

We present here the theory of Lazarus [135] and the theory of Frijda [81] which we will take as foundation models to formalize emotions at the formal level.

1.3.1 The theory of Lazarus

Lazarus [135] argues that the final goal of our emotions is to help the organism survive in the environment. His theory is a "relational" theory of emotions, in that it assumes that emotions arise from an adaptation of the individual to the environment. Stable relationships between the individual and the environment result in recurrent emotional patterns in the individual. Emotion is due to an evaluation of the potential consequences of a situation for the individual.

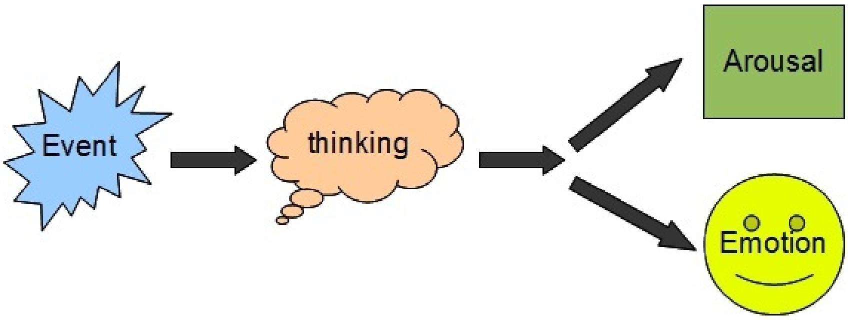


Figure 1.4: The sequence leading to emotion in the theory of Lazarus

The sequence thus is as follows (Figure 1.4): event $\rightarrow$ thinking $\rightarrow$ simultaneous arousal and emotion. Once the event occurs, after having thought, both emotion and arousal simultaneously occur.

There are two levels of appraisal. In *primary appraisal*, we consider how the situation affects our personal well-being. In *secondary appraisal* we consider how we might cope with the situation.

Primary appraisal assesses the relevance of the encounter to the individual's well-being. It has three components:

- *Goal relevance* is the importance of the situation for the individual: does the situation involve issues about which an individual cares? If the situation is not relevant to any goal, then it can trigger no emotion.

- *Goal congruence* if the situation is congruent with one of the individual's goals: if it facilitates achievement, then a positive emotion will be triggered; if it is incongruent with some goal, if it threatens or impedes its achievement, then a negative emotion will be triggered.

- Type of *ego-involvement* gathers several features of ego-identity and personal commitments, sorted in six categories: self- and social esteem, moral values, ego-ideals, meanings and ideas, well-being of other persons, and life goals. It represents in which way the agent is personally involved in the current situation.

Secondary appraisal assesses the individual's coping options, such as the actions s/he may perform and her/his envisaged effects on the situation. This appraisal level has three components:

- *Blame or credit* are an attribution of responsibility to a person accountable for the situation at hand. If this person had control on what happened, then s/he can receive blame or credit. Did someone deliberately provoke this situation?
- *Coping potential* is an evaluation of how the individual can manage the situation, change it or change her/his goals, in order to restore a good relationship with her/his environment. Can s/he do something to restore the balance between her/him and environment?
- *Future expectations* represent the expected modifications of the situation if the agent does not intervene: will the situation turn out right if the individual does nothing?

Each type of emotion can be defined by a relational meaning which expresses the set of benefits and harms to an individual in some environment, constructed by a process of appraisal. Each type of emotion is distinguished by a pattern of appraisal factors. Lazarus identified the *pattern of appraisal factors* for nine negative emotions, including *anger, fright, anxiety, guilt, shame, sadness, envy, jealousy, disgust*; and six positive emotions, including *joy or happiness, pride, love or affection, relief, hope and compassion.* This pattern of appraisal factors enables to develop a model of emotion for computer application.

1.3.2 The theory of Frijda

Frijda [81] points out that the word "emotion" does not refer to a "natural class" and that it is not able to refer to a well-defined class of phenomena which are clearly distinguishable from other mental and behaviour events.

At the center of Frijda's theory is the term *concern*. A concern is the disposition of a system to prefer certain states of the environment and of the organism, over the absence of such conditions. Concerns produce goals and preferences. If the system has problems to realize these concerns, emotions develop.

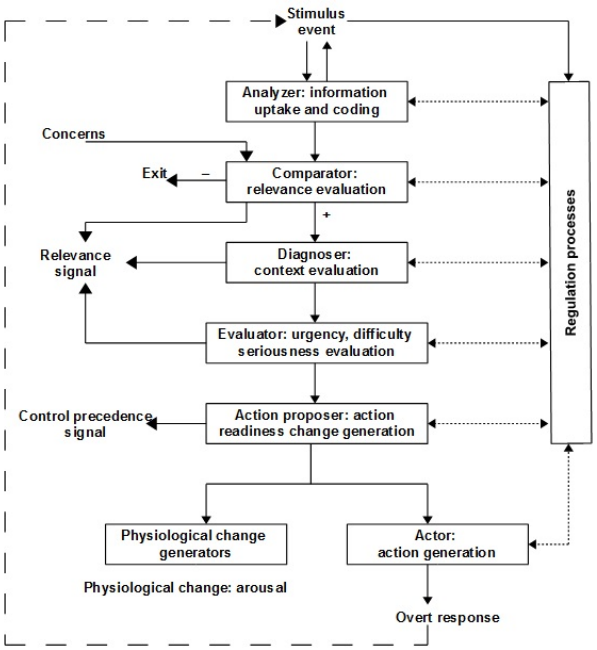


Figure 1.5: Frijda's emotion system [81]

According to Frijda, the core of the emotion process is composed of seven steps (components), as described in Figure 1.5:

- Analyser: Observation of the incoming information and subsequent coding of its implications and consequences.
- Comparator: Test of whether information is of relevance to a concern. The result are relevance signals, which activate the action system and the Diagnoser and cause attentional arousal.
- Diagnoser: Responsible for context evaluation, scanning the information for action-relevant references. Performs a number of tests (e.g.

whether consequences of an event are safe or uncertain, who is responsible for it etc.) and results in an appraisal profile.

- Evaluator: Agreement or discrepancy signals of the Comparator and the profile of the Diagnoser are combined into the final relevance signal and its intensity parameter. The intensity signals the urgency of an action to the action system. The relevance signal constitutes the so-called control precedence signal.

- Action Proposer: Prepares the action by selecting a suitable alternative course of action and by making available the resources necessary for it.

- Physiological Change Generator: effects the physiological change in accordance to the action readiness mode generated before.

- Actor: Generates actions.

Frijda also argues that cognition is a determinant of emotional response, through processes of "appraisal", "interpretation" or "meaning analysis"; and that it is a constituent of emotional experience. He thus proposes the situational meaning structures of emotion: "Each emotion corresponds to a different appraisal - a different situational meaning structure - and is characterized by it" [81, p.195]. So Frijda argues that emotion experience can take three different forms: awareness of situational meaning structure, awareness of autonomic arousal, and awareness of action readiness.

Frijda accordingly defines the situational meaning structures of *fear, anxiety, anger, sadness and grief, joy, guilt and remorse.*

1.4 Cognitive theories

The main idea of cognitive theories is that emotions are mental states which can be recognized and distinguished, based on their cognitive decompositions or propositional content. While appraisal theorists generally allow that the cognitive processes underlying emotion can be either conscious or unconscious, and can involve either propositional or non-propositional content, cognitivists typically claim that emotions involve propositional attitudes (cf. de Sousa [235]). Many emotions are specified in terms of propositions: one can not be angry at someone unless one believes that person guilty of some offense; one can not be envious unless one believes that someone else has something good in her possession. Proponents of cognitivism universalize this feature of certain emotions, and maintain that in order to have an emotion, one must always have some sort of attitude directed at a proposition.

Cognitivists claim that emotions depend upon our opinion as to what is true, not on what is in fact true. All emotions depend upon our cognitive states about the world, not the world itself. The primary test as to whether something is an emotion of the belief based sort is to ask this question: If the belief that a preference was satisfied or not were to change, would the feeling go away? If "yes", then feeling is an emotion. The major test as to whether something is an emotion of the more general sort is found by asking this question: Is it possible that a change in a belief or attitude would make this feeling go away? If "yes", then feeling is an emotion (cf. Tomlinson [245]).

The most parsimonious type of cognitive theory follows the Stoics in identifying emotions with judgments. Robert Solomon [234], Jerome Neu [175] and Martha Nussbaum [177] take this approach. Our anger at someone simply is the judgment that we have been wronged by that person. Other cognitive theories introduce further elements into their analyses. Emotions have been described as sets of beliefs and desires (Marks [151]), affect-laden judgments (Broad [30] and Lyons [147]), and as complexes of beliefs, desires, and feelings (Oakley [178]).

We present here two theories in this category as some particular examples: the theory of Ortoney et al. [186], and the theory of Reisenzein [205].

1.4.1 The theory of Ortony, Clore and Collins

Ortony, Clore and Collins [186] developed their theoretical approach with the express aim to implement it in a computer. They assume that emotions develop as a consequence of certain cognitions and interpretations. Therefore the theory exclusively concentrates on cognitive elicitors of emotions, including *events*, *agents* and *objects*.

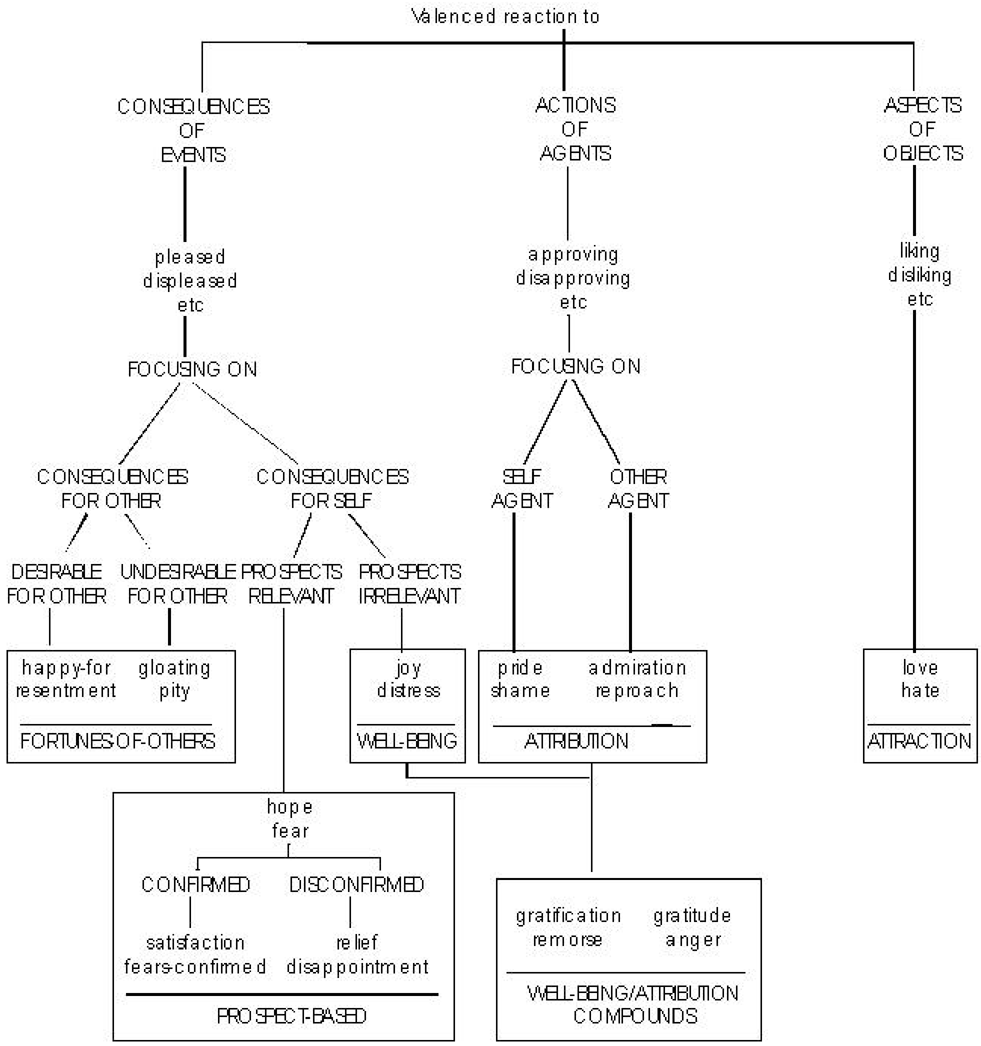


Figure 1.6: Structure of emotion types in the theory of Ortony, Clore and Collins [186]

Each cognitive elicitor has some local variable such as *desirability* for an event or for other, *confirmation* or *dis-confirmation* of an event, *praisewor-*

thiness or *blameworthiness* of an agent's action, *appealingness* for an object. Each emotion has a unique combination of these variables' values, which is called the cognitive structure of such an emotion. All possible combinations yield twenty-two types of emotions as in Figure 1.6.

Reaction to events There are four groups of emotions:

- Well-being emotions (joy, distress) correspond to the appraisal of the desirability for the self of an event that has occurred.
- Prospect-based emotions (hope, fear) correspond to the appraisal of the desirability for the self of an event that may occur.
- Confirmation emotions (satisfaction, fear-confirmed, disappointment, relief) correspond to the appraisal of the desirability for the self of an event that was expected before it occurred (these emotions thus arise after a corresponding prospect-based emotion).
- Fortunes-of-others emotions (happy-for, sorry-for, gloating, resentment) correspond to the appraisal of the presumed desirability for another agent of an event.

Reaction to agents

This includes attribution emotions (pride, shame, admiration, reproach), corresponding to the appraisal of the praiseworthiness or blameworthiness of an agent's action.

Reaction to objects

This includes attraction emotions (love, hate), corresponding to the appraisal of the appealingness of the aspects of an object.

Compound emotions

This includes emotions corresponding to simultaneous appraisals of the situation, and depends upon two variables: desirability and praiseworthiness. Well-being and attribution composed emotions (gratification, remorse, anger, gratitude) correspond to the appraisal of the consequences of an action and of the praiseworthiness of its author.

1.4.2 The theory of Reisenzein

Reisenzein [205] introduced the Belief-Desire Theory of Emotion (BDTE) based on the belief - desire theory of action [28] approach to artificial agents [27]. Posited emotion-producing mechanisms are analogous to sensory transducers; however, instead of sensing the world, they sense the state of the belief - desire system and signal important changes in this system, in particular the fulfillment/frustration of desires and the confirmation/disconfirmation of beliefs.

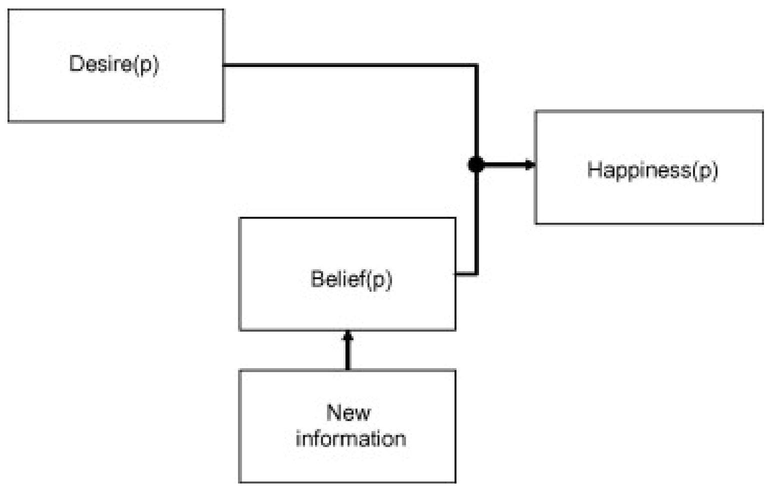


Figure 1.7: Basic belief-desire analysis of emotions in theory BDTE [205]

The process of emotion generation, illustrated for happiness, typically looks as shown in Figure 1.7. First, the person comes to desire some state of affairs or proposition p. Some time later - as a result of new information acquired through the senses, communication from others, or inference from existing beliefs - the person acquires the belief that p obtains. Thereupon, the emotion of *happiness* occurs.

BDTE analyses emotions on two aspects: qualitative (as described in Table 1.1) and quantitative for seven emotions: *happy, unhappy, hopes,*

Emotion	Belief at t	Desire at t	Belief at $t-1$
Happy(p, t)	Certain(p, t)	Desire(p, t)	
Unhappy(p, t)	Certain(p, t)	Desire$(\neg p, t)$	
Hopes(p, t)	Uncertain(p, t)	Desire(p, t)	
Fears(p, t)	Uncertain(p, t)	Desire$(\neg p, t)$	
Surprised(p, t)	Certain(p, t)	-$(irrelenent)$	Bel$(\neg p, t-1)$
Disappointed$(\neg p, t)$	Certain(p, t)	Desire$(\neg p, t)$	Bel$(p, t-1)$
Relieved$(\neg p, t)$	Certain$(\neg p, t)$	Desire$(\neg p, t)$	Bel$(\neg p, t-1)$

Table 1.1: Qualitative analysis of theory BDTE [205]

fears, surprised, disappointed and relieved. There are two main variables in this model:

- Desire: $Desire(p, t)$ reads "agent desires that p at the time t"
- Belief: is represented by three operators. $Bel(p, t-1)$ reads "agent believed p was true at the previous instant". $Certain(p, t)$ reads "agent is sure that p is true at the time t". And $Uncertain(p, t)$ reads "agent is not sure about p at the time t".

The qualitative analysis is useful for constructing a formal model of emotion, while the quantitative analysis is useful for constructing a computational model of emotion.

1.5 Emotion and Mood

Although the words "emotion" and "mood" are frequently used interchangeably, psychologists agree that the constructs they represent are closely related but distinct phenomena (cf. Beedie et al. [17]). Because our work is concerned with emotions and mood, we briefly summarize below how these two concepts differ from each other.

Cause. Moods are general, background states of feeling, with no specific cause. Emotions have a specific cause. Ekman [66] argued that people can usually specify the event that called forth an emotion, and often cannot do so for a mood. Similarly, Brehm [29] suggested that while emotions result from specific instigators, moods can occur without apparent cause.

Duration. Moods presumably endure longer than emotions. Parkinson et al. [190] proposed that emotions usually seem to have a clear moment of onset then dissipate fairly rapidly, whereas moods often change more slowly and continue to linger somewhere in the background of consciousness. However, others proposed that emotions endure longer than moods. Lazarus [136] in particular argued that the duration of emotion and mood is not a reliable criterion by which to distinguish the two constructs.

Consequences. Davidson [55] argued that mood biases cognition, whereas emotion biases behavior; Frijda [80] argued that emotions alter action readiness whereas moods produce generalized cognitive consequences; and Oatley and Jenkins [179] argued that emotions serve to rearrange the priorities of goals and change the flow of action, whereas moods maintain a distinctive readiness that continues despite events that might disturb it.

Intentionality. Emotions are always about, or directed at, something (i.e., they are intentional) while moods may not be. Frijda [80] argued that whereas one is angry (an emotion) about or at something, we tend to be irritable (a mood) non-specifically. Parkinson et al. [190] also suggested that whilst moods are unfocused, emotions are directed at specific objects.

Intensity. Moods are generally less intense than emotions. Mandler [150], for example, described mood as a persisting state of low level emotion; Lang [132] described moods, in relation to emotions, as less intense but more persistent states of feeling; and Panksepp [189] referred to the arousal associated with mood as milder and more sustained than that associated with emotion.

Function. Some author suggested that mood signals the state of the self whereas emotion signals the state of the world (Frijda [80], Morris [167], Schwarz and Clore [226]). Lazarus [136] similarly argued that emotions refer to the immediate adaptational business in an encounter with the environment, the fate of a specific and narrow goal that confronts a beneficial or harmful (or threatening) environmental condition whilst moods are products of appraisals of the existential background of our lives.

1.6 Conclusion

This chapter gave a selective overview of research on emotion in psychology, including three main categories: feeling theories which say that emotions are triggered by the perception of the physiological changes in the body; appraisal theories which consider that emotions involve evaluations about related events which are favorable (good) or harmful (bad) to the owner interests; and cognitive theories which say that emotions are mental states which can be recognized and distinguished, based on their cognitive decompositions or propositional content. We investigate into appraisal and cognitive theories as foundation models in order to formalize emotion at a formal level, which will be presented in chapter 5.

The next chapter will present researches on emotion for computer science, particularly the processing of emotions in ambient intelligent systems and affective computing.

Chapter 2

Emotion in Computer Science

This chapter presents a selective overview of researches in emotion for applications in computer science. Section 2.1 presents the reasons why we need emotional processing in computer application; Section 2.2 introduces the affective computing; Section 2.3 presents the processing of emotion in affective computing.

2.1 Need of emotions in computer science

Recently, emotions play a more and more important role in intelligent behavior researching in Artificial Intelligence (AI). There are some motivations for reason why emotions become an active research area in computer science.

First, having emotions abilities enable agents more believable. The AI researchers try to create the illusion of life by endowing agents with human abilities that they consider essential in intelligence (like reasoning, problem solving, learning... cf. Bates [15]). Other researchers show this central role of emotions and personality in believable agents (Rousseau and Hayes-Roth [213]). Gratch and Marsella [96] agree on the crucial role emotions play in human intelligence and on the need to integrate them into virtual agents. However they notice that computer scientists are mainly interested in making their agents more believable by endowing them with emotional expressions. The central role of emotion in the design of intelligent virtual agents is thus well recognized now: emotion makes these agents more believable, and their interaction with the user.

Second, emotions have an impact on cognition. Psychologists also propose models of the impact of emotions on cognitions. Forgas [79] proposes the Affect Infusion Model (AIM), a framework allowing to explain how affectively loaded information can modify reasoning. In his model, Forgas proposes four strategies of information processing: direct processing retrieves a reaction already triggered in the past; motivated processing is a selective research, directed by a motivation towards a precise target; heuristic processing uses a limited set of information and associations to produce a low-cost answer when the two first strategies do not work; substantial processing is only used for complex new tasks. Numerous researches attest the impact of emotions on abilities characteristic of human intelligence (planning, memory, learning, etc. cf. Gratch and Marsella [96]).

Third, that is because of the birth of affective computing whose objective is to construct a transparent and adaptive interface between human and machine. In order to do this, the computer must focus not only on user's input, but also on the user's emotional states. This gives computers the human-like abilities of observation, interpretation and generation of affect features, including user's emotions. The ability of interpretation seem to be at the core of emotional processing because it requires the technologies which enable computers to understand, represent and reason about emotions.

As a result of the growing number of proofs of the impact of emotions on interaction, Picard [195] assumes that some computer systems may need to be able to recognize the user's emotions. She believes that affective computing improves interfaces by allowing affective communication with the user, and providing ways to exploit the received affective information. Mainly, interfaces can reduce the user's frustration by helping him express his emotions and then recognizing his emotional expressions. Picard highlights that it is important to let the user choose if he wants to express his emotion or not: the user wants to keep control over his expression.

The concept of affective computing and its potential applications will be presented in section 2.2; and the treatment of emotion in affective computing will be presented in detail in section 2.3.

2.2 Affective computing

2.2.1 Concept of affective computing

Affective computing is a branch of the study and development of artificial intelligence that deals with the design of systems and devices that can recognize, interpret, and process human emotions. In other terms, affective computing tries to assign computers the human-like capabilities of observation, interpretation and generation of affect features. It is an important topic for the harmonious human-computer interaction, by increasing the quality of human-computer communication and improving the intelligence of the computer.

Affective computing concerns multidisciplinary knowledge background such as psychology, cognitive, physiology and computer sciences. Main contributions of affective computing technologies are as follow:

- Development new sensors/algorithms that can detect affective states
- New techniques to infer a person's affective or cognitive state
- Developing machines that can respond affectively to a person's affective state
- Inventing personal technologies that improve machine awareness and communication

2.2.2 Potential applications

The potential applications of affective computing is mainly lied on improving the Human-Machine Interaction. We here list some kind of these interactions.

Ambient intelligent systems. In computing, ambient intelligence (AmI) refers to electronic environments that are sensitive and responsive to users. Ambient intelligence is a vision on the future of consumer electronics, telecommunications and computing. In an ambient intelligence world, devices work in concert to support people in carrying out their everyday life activities, tasks and rituals in easy, natural way using information and intelligence that is hidden in the network connecting these devices. For instance, the Oxygen project [188] aims at human-centered communication and computing through a combination of perceptual interaction, individualized knowledge access, software agents, and collaboration technologies. The Ambience project [116] focuses on networked context-aware environments. The goal of the Ozone project [115] is to investigate, define, and implement a generic framework to enable consumer-oriented ambient intelligence applications. The Digital Living Network Alliance (DLNA) [60] project aims to create a wireless inter-operable network of personal computers, consumer electronics and mobile devices in the home enabling a seamless environment for sharing and growing new digital media and content services. The AmE Framework [267] is the framework for emotion-aware ambient intelligence and present the overall structure of emotion in English conversation. Other well known projects include Gaia [207], One.world [97], and Aura [84].

The ambient intelligence paradigm is not a pure research domain, it is composed of several research domains as well as technologies which could be grouped into three mains areas: Ubiquitous computing, ubiquitous communication, and intelligent user interface which focuses on the interaction between "ordinary people" and ambient agents. In particular, the Human-Machine Interaction which is the core of affective computing.

E-learning. Affective computing can be used to adjust the presentation style of a computerized tutor when a learner is bored, interested, frustrated, or pleased. Objective of e-learning is establish a conversation between agent and studier as if it is a real social interaction between humans. Asteriadis et al. [11] present a mechanism which compiles feedback related to the behavioral state of the user (e.g. level of interest) in the context of reading an electronic document. They proposed scheme is tested in an e-learning environment, in order to adapt the presentation of the content to the user profile and current behavioral state. D'Mello et al. [61] develop a version of AutoTutor [91] that is sensitive to both the cognitive and affective states of the learner. Such an affect-sensitive tutor would presumably enhance the intelligent learning environment. Heylen et al. [106] study the ways to have the tutor agent make informed decisions based on what the agent can assume about the mental state of the student. Katsionis and Virvou [128] describe how system observations of students' behavioural characteristics, during their interaction with an educational application, may provide important evidence about students' emotions while they learn.

Games. In game, there is a demand for the personalities, moods, and relationships of Non Player Characters' (NPCs) to be made the focus of game-play. MacNamee and Cunningham [148] present the μ-SIC system, a component of an intelligent agent architecture designed for the creation of NPCs for computer games. μ-SIC uses a number of quantitative psychological models to simulate characters' personalities, moods and relationships. The values of these models are used as inputs to an artificial neural network which drives characters' social behaviour.

Psychological health services. These services concern counseling, benefit from affective computing applications when determining a client's emotional state. For instance, Lisetti and LeRouge [144] introduce an intelligent interface aimed at improving affective distance communications between a patient and health care provider and patient emotional state monitoring in the tele-home healthcare context.

Robotic systems. They are capable of processing affective information exhibit higher flexibility while one works in uncertain or complex environments. Affective computing could offer more natural interfacing and more engaging interaction between human-robot or between robots. For instance, Jones and Deeming [122] reports on the development and evaluation of a consumer-level robotic dog with acoustic emotion recognition capabilities. The dog can recognise the emotional state of it's owner from affective cues in the owner's speech and respond with appropriate actions.

Decision making and planing. Decision making and planning depend not only upon the internal state and the goal of agent, but also upon the environment within agent is situated and for which plans are generated. It is the reason why we need to take into account the context, or the interaction between an autonomous situated planning agent and its environment. For instance, Coddington and Luck [48] describe how context constrains the goals an agent might generate, enables those goals to be prioritised, and constrains plan selection. Marsella and Gratch [152] address how emotions arise from an evaluation of the relationship between environmental events and an agent's plans and goals, as well as the impact of emotions on behavior, in particular the impact on the physical expressions of emotional state through suitable choice of gestures and body language. They also designed their model to support characters that act in virtual environments, make decisions and planning.

Other potential applications are centered around social monitoring. For example, a car can monitor the emotion of all occupants and engage in additional safety measures, such as alerting other vehicles if it detects the driver to be angry. In other human-computer interaction, affective mirrors allow the user to see how he or she performs; emotion monitoring agents sending a warning before one sends an angry email; or even music players selecting tracks based on mood.

2.3 Emotion processing in affective computing

Generally, emotion is treated in affective computing by two main phases. Firstly, detect user's emotion from their vocal, facial or body signals. Secondly, representing them in the way that the computer can understand, calculate and reason about. This section presents the researches in each category.

2.3.1 Speech processing

For emotional speech processing, it is a widely known fact that the emotional speech differs with respect to the acoustic features [221]. The acoustic features are therefore widely used for the research of emotion recognition with the pattern recognition methods. For instance, Scherer [220] performed a large-scale study using 14 professional actors. Dellaert [57] used prosody features and compared three classifiers: the maximum likelihood Bayes classification, kernel regression, and k-nearest neighbor in emotion recognition for sadness, anger, happiness and fear. Lee [137] used linear discriminant classification with Gaussian class-conditional probability distribution and k-nearest neighborhood methods to classify utterances into two basic emotion states, negative and non-negative. Yu [265] used support vector machines for emotion detection. Nick [34] proposed a perception model of affective speech utterances. Williams and Stevens [258] studied the spectrograms of real emotional speech and compared them with acted speech. Murray and Arnott [171] reviewed findings on human vocal emotions. They also constructed a synthesis-by-rule system to incorporate emotions in synthetic speech [172]. Chiu et al. [44] extracted five features from speech and used a multi-layered neural network for the classification. Petrushin [194] compared human and machine recognition of emotions in speech and achieved similar rates for both. Chen [43] proposed a rule-based method for classification of input audio data into one of the following emotions categories: happiness, sadness, fear, anger, surprise, and dislike.

For emotion generating with speech synthesis, Mozziconacci and Hermes [169] added emotion control parameters on the basis of tune methods resulting in higher performance power of voice composing. Cahn [63], by means of a visualized acoustic parameters editor, achieved the output of emotional speech with manual inferences. Nick [35] created an expressive speech synthesis with a five years' large corpus and gave us an impressive synthesis results. Schroeder and Breuer [225], Eide et al. [65] generated a expressive text-to-speech (TTS) engine which can be directed, via an extended Speech Synthesis Markup Language, to use a variety of expressive styles with about ten hours of "neutral" sentences. Chuang and Wu [46] and Tao [242] used emotional keywords and emotion trigger words to generate the emotional TTS system.

There are currently two main challenges for speech processing. Firstly, the lack on the capture and the analysis of more detailed/reliable physiological features limits the further improvement of the research. Secondly, people express the feeling not only by the acoustic features, but also with the content they want to say. Different words, phrases and syntactic structures, etc. can make many kinds of expression results and styles. Moreover, the acoustic features are different in different languages.

2.3.2 Facial expression

Facial expressions and movements such as a smile or a nod are used either to fulfill a semantic function, to communicate emotions, or as conversational cues. For instance, Etcoff and Magee [74] parameterized the structure of the chief parts of human's face through 37 lines, which enables people to roughly tell the affect status of faces. Paul Ekman [67] and his colleagues have performed extensive studies of human facial expressions. They found evidence to support universality in facial expressions. These "universal facial expressions" are those representing happiness, sadness, anger, fear, surprise, and disgust. Ekman and Friesen [68] developed the Facial Action Coding System to code facial expressions where movements on the face are described by a set of action units. Pelachaud et al. [155, 176, 193] created Greta, an affective Embodied Conversational Agent (ECA), that is able to display communicative and emotional signals. Greta can talk and simultaneously show facial expressions, gestures, gaze, and head movements.

To do the facial expression analysis, most of the facial features were captured by the optical flow or active appearance model. Mase [156] used optical flow to recognize facial expressions. He was one of the first to use image processing techniques to recognize facial expressions. Lanitis et al. [133] used a flexible shape and appearance model for image coding, person identification, pose recovery, gender recognition and facial expression recognition. Black and Yacoob [20] used local parameterized models of image motion to recover non-rigid motion. Once recovered, these parameters are fed to a rule-based classifier to recognize the six basic facial expressions. Yacoob and Davis [260] computed optical flow and used similar rules to classify the six facial expressions. Rosenblum et al. [211] also computed optical flow of regions on the face, then applied a radial basis function network to classify expressions. Essa and Pentland [73] also used an optical flow region-based method to recognize expressions. Otsuka and Ohya [187] first computed optical flow, then computed their 2D Fourier transform coefficients, which were then used as feature vectors for a hidden Markov model (HMM) to classify expressions. Chen [43] used a suite of static classifiers to recognize facial expressions, reporting on both person-dependent and person-independent results. Cohen et al. [49] describe classification schemes for facial expression recognition in two types of settings: *dynamic classification* by learning the structure of Bayesian networks classifiers; and *static classification* by using a multi-level HMM classifier. Lyons et al. [146] applied the supervised Fisher linear discriminant analysis to learn an adequate linear subspace from class-specified training samples and the samples projected to this subspace can be best separated.

Another fact of facial processing is body gesture and movement, which is defined by the positions of body arthroses and their changes with time. Currently, the work for gesture processing is more focused on the hand tracking. Hand gestures can convery various and diverse meanings, to enhance the mood or to behave as a sign language. For instance, Pavlovic et al. [192] used apparentness methods; while Aggarwal and Cai [5] used 3-D modeling methods.

When the interaction is multimodal in which participants encounter a steady stream of meaningful facial expressions, gestures, body postures, head movements, words, grammatical constructions, and prosodic contours, we need a multimodal processing of emotion. For instance, in biometrics recognition systems, Brunelli and Falavigna [31] describe a multimodal biometric system that uses the face and voice traits of an individual for identification. Bigun et al. [19] develop a statistical framework based on Bayesian statistics to integrate information presented by the speech (text-dependent) and face data of a user. Kumar et al. [130] combined hand geometry and Palmprint biometrics in a verification system. A commercial product called BioID [82] uses voice, lip motion and face features of a user to verify identity. Jain and Ross [117] improved the performance of a multimodal system by learning user-specific parameters. General strategies for combining multiple classifiers have been suggested in Ho et al. [109] and Kittler et al. [129]. There is a large amount of literature available on the various combination strategies for fusing multiple modalities using the matching scores.

Recently, a new approach for audio-visual mapping has arisen [51], which is inspired from speech synthesis [112]. This method means to construct new data stream by concatenating stored data units in training database. It has advantage of that the synthesis result appears very natural and realistic. But even for that, the lip movement is still the focus in most of the research. Full facial expression, especially the correlation between facial expression and more acoustic features, such as prosody, timbre, etc. has seldom been touched. It needs more work in processing the features which we ignored before.

2.3.3 Affective understanding and reasoning

After having recognized, emotion will be entered into next steps of emotional processing, including representing, calculating or reasoning about emotion.

2.3.3.1 Representation of emotion

Objective is find out the way to represent emotion in order to make understand to the computer. Thus almost current models are based on a basic cognitive model of emotion in psychology as a foundation. For instance, Yao et al. [261] propose an agent model of emotion development. The agent model is based on the cognitive model of Ortony et al. [186] and Operant Conditioning Theory [168, 138]. Oliveira et al. [183] present an agent architecture that includes several emotional-like mechanisms, namely: emotional evaluation functions, emotion biased processing, emotional tagging and mood congruent memory. Van Dyke Parunak et al. [191] present DETT (Disposition, Emotion, Trigger, Tendency), a model of emotion for situated agents that captures the essential features of the Ortony et al. [186] model in a computationally efficient framework. DETT combines the theoretical richness of Gratch-Marsella with the computational efficiency of EINSTein and MANA [95], and adds dispositions, accounting for differences in the emotional susceptibility of various agents by reasoning about the reaction of an internal agent characteristic to the external environment in which the agent is situated. Stephane [237] presents the integration of cognitive and emotional models that comprise on one hand users' Situation Awareness and on the other hand users' Self Awareness. This framework is integrated from two models: the Schematic, Propositional, Analogical and Associative Representation Systems (SPAARS) [199] and the Self-Regulatory Executive Function (SREF) [256].

2.3.3.2 Computation of emotion

Objective is to qualitative emotion, in some concrete context, based on the calculation of its cognitive factors.

Gershenson [87] proposes a model for emotions, which handles the degree of contradiction that emotions might have. This model is based on the multidimensional logic, which is a paraconsistent logic [200], and the cognitive structure of emotions proposed by Ortony et al. [186]. They represent an emotion by a tuple of three dimensions: the x axis represents positive emotions, the y axis negative ones, and they order their specific emotions in the z axis arbitrarily. The specific emotions are emotions they consider as the basic pair of emotions: *love/hate*, *joy/grief*, *happy/sadness*. So, in the z axis, the zero represents *love/hate*, the one *joy/grief*, the two *happy/sadness*, and the three *love/hate* again. This makes the z axis cyclic. So an emotion is represented by an vector of three dimensions, for example, *pride* is presented by vector $near(1, 0, 1.5)$. It means that "pride is a positive emotion (positive = 1, negative = 0) with the degree of satisfaction is between joy and happy".

Hu et al. [111] present an emotional agent model based on the Belief-Desire-Intention (BDI) agent model [86]. Assuming that there is N emotion types, they use a vector space E to represent the emotions, $E =< e_i|i = 1, 2, ..., N >$, where N represents the number of the emotion types. For the fuzzy intensity of emotion they use a discrete variable V_i and divide it into M levels. Then emotion model can be represented simply as:

$$\begin{pmatrix} E \\ V \end{pmatrix} = \begin{Bmatrix} e_1 & e_2 & ... & e_n \\ v_1 & v_2 & ... & v_n \end{Bmatrix}$$

Once the types of emotion is fixed, the emotion model can be simplified as a vector $E =< v_1, v_2, ..., v_n >$. In the light of the psychology research they divide the emotion into five basic types: *anger, afraid, happy, sad, dislike*. For instance, $E =< 0.25, 0.5, 0, 0, 0 >$ represents the autonomic unit which has the complex emotion of a *little angry* also *very fear*.

Steunebrink et al. [238, 239] treat the questions of which quantitative aspects of emotions (of the model of Ortony et al. [186]) should be considered, how these quantitative aspects can be formally modeled, and how a quantitative model of emotions can be combined with the qualitative model. Quantitative aspects of emotions are described in terms of *potentials*, *thresholds*, and *intensities*. They construct three functions for each of twenty two emotion types in Ortony et al. [186] model.

For instance, with Joy state, the potential function P_{joy} is defined as the weighted sum (denoted as $\overrightarrow{w}$) of the variables (denoted as $\overrightarrow{x}$) that according to Ortony et al.'s affect intensity, $P_{joy}(a, s) := \overrightarrow{w}\Delta\overrightarrow{x}$. The threshold function T_{joy} is defined as: $T_{joy}(a, s) := \overrightarrow{w}\Delta Mood(a, s)$, where $Mood(a, s)$ depends on the Ortony et al.'s "mood" of the agent. And the intensity functions I_{joy} is defined as the difference between the potential and threshold as the numerator (if this difference is negative, the intensity should be set to zero altogether), and the difference between the current time and triggering time in the exponent: $I_{joy}(a, s', t') := max(0, \frac{q_p - q_t}{1+e^{(t'-t-\mu)\delta}} - \theta).$, where the half-life μ, fall-off speed δ, and cut-off threshold θ, such that $\mu\delta \approx -ln0.01$ and $0 < \theta \ll q_p - q_t$.

2.3.3.3 Reasoning about emotion

Objective is to qualitative emotion and represent emotion in a formal form in order to enable the computer to understand and reason on the emotions.

Sanders [218] has defined a logic in which common sense knowledge about human psychology can be expressed. They focus on a cluster of emotions, including *approval, disapproval, guilt and anger*, most of which involve some sort of ethical evaluation of the action that triggers them. They use an extension of temporal logic developed by Shoham [231], modified to incorporate the three modal operators "know", "believe" and "want": they use an S5 (cf. Chellas [42]) axiom set for "know", weak S5 for "believe" and T for "want". They also use some basic concepts of deontic logic [160] such as obligation, prohibition and permission.

Meyer [164] presented a logic which enables to reason about the dynamics of emotional states of agent. This logic is based on the BDI model [86], the framework of dynamic logic and (an extension of) the KARO framework [249]. They provide KARO-style formulas expressing emergence of the four basic emotions of *happiness, sadness, anger* and *fear*, as well as their influence on the agent's deliberative behavior. For instance, the formula $I(\pi, \varphi) \wedge Com(\pi) \to [\pi](\varphi \to happy(\epsilon, \varphi, \varphi))$ expresses that an agent that was committed to a plan π which it intended to do with φ as goal, and that believes that its goal φ is realized after having executed/performed its plan π, is indeed happy with this (w.r.t. this goal φ and a empty plan ϵ). In this formulas, $I(\pi, \varphi)$ means that agent intends to do action π to achieve φ; and $Com(\pi)$ means that π is committed to do by such an agent.

Ochs et al. [181, 182] presented a computational model of capability-based emotion elicitation for rational agent, this model is also based on the cognitive appraisal theory of Ortony et al. [186]. The mental state of a rational agent is composed of two primitive mental attitudes: *belief* and *choice*, which are formalized with the modal operators B and C.

According to cognitive appraisal theory of emotions [186], capability-based emotions are triggered by events that are evaluated as disturbing beliefs on one own capabilities, they consider that an occurred event is evaluated as *praiseworthiness* if one of agent's capabilities has been surpassed.

So, *pride* elicitation is formalized as [181]:
$B_i(p) \wedge Done_i(e, B_i(\neg Can_i(e,p)) \wedge C_i(p)) \Rightarrow Pride_i(e,p)$
which reads that agent i triggers pride emotion after the event e if she has achieved p (p being a closed formula denoted a proposition) by e whereas she thought it was not possible to do. In which, $B_i(p)$ means "agent i believes that p is true". $C_i(p)$ means "agent i chooses/desires that p is currently true". $Can_i(e,p)$ means that agent i has the capability to perform the event (or action) e and, consequently, p which is an effect of this event.

Adam et al. [3] provide a logical formalization of twenty emotions while staying as closed as possible to the model of Ortony et al. [186]. These emotions are formalized inside a BDI modal logic [86]. Their formal framework is based on the modal logic of belief, choice, time, and action of Herzig and Longin [105] which is a refinement of Cohen and Levesque's works [50]. For instance, *satisfaction* is formalized as:
$Satisfaction_i\varphi \overset{def}{=} Bel_i P Expect_i\varphi \wedge Des_i\varphi \wedge Bel_i\varphi$
where $Satisfaction_i\varphi$ means that agent i is satisfactory about φ (a proposition); $Bel_i\varphi$ means that agent i believes that φ is true; $Des_i\varphi$ means that agent i desires that φ is the case; $P\varphi$ means that φ was true in the past; and $Expect_i\varphi \overset{def}{=} Prob_i\varphi \wedge \neg Bel_i\varphi$ means that agent i expects φ to be true but envisages the possibility that it could be false.

Basically, the logical framework presented in this thesis is also based on the logic of *belief* and *choice*, as the logic of Adam et al. [3] used. However, there will be some extents in order to better represent emotions and trust/distrust. Detailed discussion will be given in section 5.1.6.

2.4 Conclusion

This chapter presented a a selective overview of researches on emotions in computer science. Particularly, we examined the research on the treatment of emotion in affective computing where users' emotion, once being recognized, could be represented, measured and reasoned in computer application.

In this thesis, we focus on the latest category, representing agent by one qualitative formalization of it for computer application based on the cognitive theories of emotion.

Chapter 3

Introduction of Trust/Distrust

In oder to give the computer the ability to reason about emotions, we aim to formalize some attitude which has a tight and natural relation with emotions: trust and distrust mental states.

In this chapter, we give a selective view of researches of trust and distrust in literature. First of all, we examine the meaning of trust and some of relevant concepts in philosophy. Then, we try to category recent models of trust, which are proposed for computer science, based on the approach they are built on. After that, we examine models in each category.

Section 3.1 represents the meaning of trust and some of concepts relevant in philosophy; Section 3.2 represents recent proposed models of trust in computer applications.

3.1 The meanings of Trust/Distrust

3.1.1 Concept of trust

Trust has been defined by researchers in many different ways. Many trust researchers have recognized and examined more than one type of trust. Some sociologists tend to see trust as structural in nature which is founded upon social or institutional structures in the situation, not on the person, or personal attributes, of the trusted parties (Lewis and Weigert [142], Shapiro [230]). Some psychologists have viewed trust as a personal attribute, such as a property or state of trusted person (Rotter [212]). Social psychologists are more likely to view trust as an interpersonal phenomenon which means that two or more people trust each other in the specific situation (Deutsch [59], Holme [110]). Economists are more inclined to view trust as a rational choice mechanism (Williamson [259]).

More general, Gambetta [83] considers trust as the subjective probability by which an individual expects another individual to perform a given action on which its welfare depends. Jøsang et al. [125, 126] consider trust as the extent to which a given party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible. Grandison and Sloman [94], based on the computational point of view, define trust as a quantified belief by a trustor with respect to the competence, honesty, security and dependability of a trustee within a specified context.

Generally, trust involves two parties, the trusted party (hereafter, called *trustee*) and the trusting party (hereafter, called *trustor*). Trustor wants or desires some goodwill that could be realized by some trustee's activity.

However, these definitions of trust are still too large and general to formalize the concept of trust for computer. We therefore need some definition which enables to recognize and distinguish trust from other related concepts: we investigate in definitions of trust from the cognitive point of view.

McNight and Chervany [162] have also defined six trust-related constructs for trust typology conceptual definitions:

- Trusting Intention: is the extent to which one party is willing to depend on the other party in a given situation with a feeling of relative security, even though negative consequences are possible.

- Trusting Behavior: is the extent to which one person voluntarily depends on another person in a specific situation with a feeling of relative security, even though negative consequences are possible.

- Trusting Beliefs: means the extent to which one believes (and feels confident in believing) that the other person is trustworthy in the situation. In which, trustworthy means one is able and willing to act in the other person's best interests.

- System Trust: means the extent to which one believes that proper impersonal structures are in place to enable one to anticipate a successful future endeavor.

- Dispositional Trust: a person has Dispositional Trust to the extent that s/he has a consistent tendency to trust across a broad spectrum of situations and persons.

- Situational Decision to Trust: means the extent to which one intends to depend on a non-specific other party in a given situation. It means that one has formed an intention to trust every time a particular situation arises, irrespective of one's beliefs about the attributes of the other party in the situation.

This definition mainly investigated in the trustor side. Trustor has to answer questions to know whether s/he will depend and being voluntary to depend on someone, whether s/he believe in someone's trustworthy, whether there is some situation in which s/he might depend on someone, etc. This seem to be a topology for pre-conditions of trust instead of a definition of trust itself.

While McNight and Chervany [162] investigate in trustor side, Jones [121] and Castelfranchi and Falcone [40, 75] investigate in the trustee side: trust depends on the belief of trustor about some trustee's attribute.

Jones [121] considers trust as a belief: it is suggested that two beliefs, called the *rule-belief* and the *conformity-belief*, form the core of the trusting attitude. Jones analyzed five scenarios in which, it is true to say that i trusts j:

- Belief in regularity of trustee: i believes that there exists a regularity in j's behavior. And in addition, i believes that this regularity will also be instantiated on some future occasion. In this case, i's rule-belief is his expectation that, in given circumstances, the event j will occur; and i's conformity-belief is his belief in regularity of j's behavior.

- Belief in trustee's obligation: i believes that there is a rule requiring j to do α, and that j's behavior will in fact comply with this rule. For instance, i believes that j is under an obligation to repay a debt, and that j will make the repayment. In this case, i's rule-belief is his belief that j is under an obligation to repay a debt; and i's conformity-belief is his belief that j will fulfill his obligation.

- Belief in role of trustee: i believes that j occupies some particular role, and that j will perform the acts associated with that role in a competent manner. For example, i trusts his doctor: i's rule-belief is his belief that, in social standard, a doctor meets the requirement of a doctor; and i's conformity-belief is his belief that j will satisfy this requirement.

- Belief in trustee's informing: i believes that j is transmitting some information to him, and that the content of j's message is reliable. In this case, i's rule-belief is his belief that j is in a convention with him; and i's conformity-belief is his belief that j will respect this convention.

- Belief in trustee's intention: i intends to see that φ is realized, and i believes that j will do nothing which will make it less likely that φ obtains. In this case, i's rule-belief is his belief that desired event will occur in given circumstance; and i's conformity-belief is his belief that j will not attempt to stop such an event.

While Jones [121] used only two kinds of belief to construct trust, this makes his definition still simple and general, Castelfranchi and Falcone [40, 75] define trust in a particular situation, in which trust is considered as a foundation condition to delegation, with some more kinds of belief.

- Only a cognitive agent, which is endowed with goals and beliefs, can trust another agent. First, one trusts another only relatively to a goal (i.e. for something s/he wants to achieve or s/he desires). If we do not potentially have goals, we can not really decide, nor care about something (welfare): We can not subjectively trust somebody. Second, trust itself consists of beliefs.

- Trust basically is a mental state, a complex attitude of an agent x towards another agent y about the behavior/action α relevant for the result (goal) g. In which, x is the relying agent, who feels trust; y is the agent or entity which is trusted. So x trusts y about g/α and for g/α; x trusts also that g will be true. Since y's action is useful to x, and x is relying on it, this means that x is delegating some action/goal in her own plan to y.

- Trust is the mental counter-part of delegation. In delegation, the delegating agent x needs (or likes) an action of the delegated agent y and includes it in her own plan: x relies on y. x plans to achieve g through y. y's delegated task is either a x's state-goal (something x wants or desires) or action-goal (something x wants or intends to do).

In other words, agent x delegates a task to agent y based on a specific set of beliefs and goals. This mental state is what they call trust. Therefore, in order to delegate his task to agent y, agent x must have trust on y. Then x has some specific beliefs:

- Competence belief: x should believe that y is useful for this goal of its, that y can produce/provide the expected result, that y can play such a role in x's plan/action, that y has some function. A positive expectation is the combination of a goal and of a belief about the future (prediction): x both believes that g and x desires/intends that g. In this case, x both believes that y can and will do; and x desires/wants that y can and will do.

- Dependence belief: x believes, to trust y and delegate to it, that either x needs it, x depends on it (strong dependence, cf. Sichman et al. [233]), or at least that it is better to x to rely than do not rely on it (weak dependence, cf. Jennings [120]).

- Disposition belief: x should think that y not only is able and can do that action/task, but also y actually will do what x needs. In case of an intentional agent, the disposition belief must be articulated in and supported by two more beliefs:

  - Willingness belief: x believes that y has decided and intends to do α. In fact for this kind of agent to do something, it must intend to do it. So trust requires modeling the trustee's mind.
  - Persistence belief: x should also believe that y is stable enough in his intentions, that has no serious conflicts about g (otherwise he might change his mind), or that y is not unpredictable by character, etc.

- Self-confidence Belief: x should also believe that y knows that he can do α. Thus he is self-confident.

- Motivation Belief: x believes that y has some motives to help her/him (to adopt her/his goal), and that these motives will probably prevail - in case of conflict - on other motives, negative for her.

- Fulfillment belief: x believes that g will be achieved (thanks to y in this case). Thus, if the agent trusts y for g, the agent decides: (i) not renouncing to goal g, (ii) not personally bringing it about, (iii) not searching for alternatives to y, and (iv) to pursue g through y.

The first two beliefs compound what they call the core trust and together with the disposition belief, the reliance.

So, after Castelfranchi and Falcone [40, 75], trust is a set of mental attitudes characterizing the delegating agent's mind x which prefers another agent y to do action α. y is a cognitive agent, so x believes that y intends to do the action and y will persist in this.

Generally, cognitivists have a common principle to construct the concept of trust: they consider trust in its decompositions (cognitive factors): the belief of trustor about a given trustee's competence, intention, integrity, power, willingness, etc. However, there is some difference among these definitions, revolving their covered scope. In particular, while Castelfranchi and Falcone [40, 75] proposed the concept of individual trust based on the trustee's attributes such as competence, power and intention. Others cover also the case, in which, the trustor trusts a trustee based on trustee's role in their institution (cf. role-based trust of Jones [121], the system trust (institution-based trust) of McNight and Chervany [162]), or based on trustee's obligation (cf. Jones [121]). In order to limit maximally the ambiguous in our latter formalization, we take the definition of Castelfranchi and Falcone [40, 75] as a foundation model of trust, on which, we formalize the concept of trust for computer, and then, find out the effect of trust on emotions.

3.1.2 The concept of distrust

Several authors agree that distrust is a partial negative form of trust, but distrust is not the negation of trust. Thus, most of them define distrust from trust's definition. For instance, Marsh and Dibben [154] argue that distrust, by comparison is a measure of how much the trustor believes that the trustee will actively work against them in a given situation. Thus, if we distrust someone, we expect s/he will work to make sure the worst (or at least not the best) will happen in a given situation.

In a cognitive point of view, McNight and Chervany [162] defined five distrust-related constructs for distrust typology conceptual definitions:

- Distrusting Intentions: means one is not willing to depend, or intends not to depend, on the other party, with a feeling of relative certainty or confidence, even though negative consequences are possible.

- Distrust-related Behavior: means that a person does not voluntarily depend on another person, with a feeling of relative certainty or confidence, when negative consequences are possible.

- Distrusting Beliefs: means the extent to which one believes, with feelings of relative certainty or confidence, that the other person does not have characteristics beneficial to one.

- Institution-based Distrust: means one believes,with feelings of relative certainty or confidence, that favorable conditions that are conducive to situational success in a risky endeavor or aspect of one's life are not in place.

- Disposition to Distrust: means the extent to which one displays a consistent tendency to not be willing to depend on general others across a broad spectrum of situations and persons.

Similarly to their topology of trust, this definition of distrust mainly investigated in trustor side. Thus the topology seem to be pre-conditions of distrust instead of distrust's attitude states.

Castelfranchi et al. [39] also define distrust as an opposite of trust: Agent i distrusts j to ensure φ by performing α if and only if:

- i wants to achieve φ; and

- i expects that
  - j has not the opportunity to ensure φ by performing action α; or
  - j is not willing (i.e., does not intend) to perform action α; or
  - the internal or external preconditions for the execution of action α by agent j do not hold.

We base mainly upon the definition of Castelfranchi et al. [39] to construct our formal definition of distrust, which will be presented in chapter 5.

3.1.3 Properties of trust

As mentioned, trust is a mental attitude of trustor which is raised from the trustor's awareness about the competence, power and willingness of trustee. Thus, trust has some properties:

- Trust is dynamic, meaning it can increase with positive experience and decrease with negative experience or over time without any experience (Ries et al. [206]).

- Trust is subjective and therefore asymmetric (Grandison and Sloman [93]): A's trust in B is not usually the same as B's trust in A.

- Transitivity of trust: If A trusts B and B trusts C, then A also trusts C? Marsh [153] points out that trust is not transitive. At least it is not transitive over arbitrary long chains, since this will end in conflicts regarding distrust. However, Grandison and Sloman [93] state that trust has an unintentional transitivity in some particular cases.

- The trustor accepts some level of risk of vulnerability (Becker [16]). The risk of vulnerability may be from the failure by the trustee to do what the truster depends on that.

- Trust has the potential for betrayal (de Sousa [236]). Annette Baier writes that trusting can be betrayed, or at least let down, and not just disappointed (Baier [12]).

- No suspicious. People do not, or cannot, trust one another if they are easily suspicious of one another (Govier [90]). Trusting involves being optimistic, rather than pessimistic, that the trustee will do something for us (or for others perhaps); and such optimism is, in part, what makes us vulnerable by trusting.

- Belief or to be optimistic of the trustee's competence. We usually trust people to do certain things - for example, to look after our children, to give us advice, or to be honest with us - but we would not do so if we thought that they lacked the relevant skills (including moral skills of knowing what it means to be honest or caring, cf. McLeod [163]).

- The commitment between truster and trustee. When we trust people, we are optimistic not only that they are competent to do what we trust them to do, but also that they are committed to doing it (de Sousa [236]).

3.1.4 Related terminologies

The following presents some concept relevant to trust: trustworthiness, reliance and reputation.

3.1.4.1 Trustworthiness

As trust, there are several points of view of trustworthiness. While both the competence and motivational elements of trustworthiness are crucial, the exact nature of the latter is unclear (de Sousa [236]). For some philosophers, it is only the ongoing commitment of the trustee (Hardin [99]). But for others, it is also the origins of the commitment matter. While some philosopher considers that trustworthiness can be "compelled by the force of norms" (Hardin [99], ONeil [184]). Some others view trustworthiness as a goodwill, which finds trustworthiness only where the trustee is motivated by goodwill (Jones [124]). However, de Sousa [236] states that goodwill is neither necessary nor sufficient for trustworthiness.

While not still agree about concept of trust and trustworthiness, philosophers agree that trust is different from trustworthiness. Trust is an attitude that we have towards people whom we hope will be trustworthy, while trustworthiness is only a property, not an attitude (de Sousa [236]).

3.1.4.2 Reliance

In one hand, trust is similar to reliance because in both concepts, there is someone who relies on other's action to obtain a goodwill. If the action is failed, both bring about disappointment. In other hand trust is also different from reliance. Baier [12] says that trust can be betrayed, or at least let down, and not just disappointed. While the latter is not. Why can trust be betrayed, but not mere reliance? The answer Baier gives is that betrayal is the appropriate response to someone on whom one relied to act out of goodwill, as opposed to ill will, selfishness, or habit bred out of indifference [12].

3.1.4.3 Reputation

Reputation is the common or general estimate of a person with respect to character or other qualities (cf. Sabater and Sierra [216]). Generally, reputation is trust in a different scope: while trust concerns an individual, reputation concerns a common group. Thus, instead of being judged by an individual truster, trustee is judged by a group of people.

In the relation between trust and reputation, in the context of a group of agents, individual trust can affect on the group trust (reputation) and vice versa, the group trust can also change the individual trust on some trustee. The mechanism to determine individual trust based on reputation is popularly applied in the computational models of trust.

3.1.5 Trust/Distrust and emotion

Trust/distrust have a tight relation with emotion. This relation has been pointed out in literature. We list here three main reasons for this relation.

First, trust/distrust is possibly considered as a particular emotion (cf. Lazarus [135, 136]). Barbalet [13] argues that trust is an emotion because of three reasons:(i) trust is feelings of positive expectation and safe dependency; (ii) trust is a feeling of confidence in another's future actions; and (iii) confidence concerning one's own judgment of another. Lahno [131] has argued that common sense is right in maintaining that trust has emotional character. In some specifiable sense, trust is an emotional attitude: they are recognized by some general pattern in the way the world or some part of the world is perceived by an individual. de Sousa [56, 236] also argues in the same direction that the characteristics of trust that suggest it is an emotion are ones that we can try to mimic in our attitude toward other people, in an effort to be more trusting. In other words, we could purposefully try to focus our attention on what makes other people trustworthy, and in doing so cultivate trust in them. In doing so, our goal might simply be self-improvement: that is, becoming more trusting, hopefully in a good way, so that we reap the benefits of justified trust.

Second, trust and distrust could be associated to some emotions. Particularly, trust is associated to hope (Lewicki et al. [140]) because they have a positive expectation (Falcon and Castelfranchi [75]), and distrust is associated to fear (Lewicki et al. [139, 140]) because they both have a negative expectation (Castelfranchi et al. [39]). As a consequence, the success or failure of trust could bring about some emotions. Particularly, the violation of trust causes some negative emotions such as anger and disappointment (Lewicki and Tomlinson [141]). This ideal is shared with Oatley [180] who argues that anger is the sentiment of antipathy and the belief of distrust. The argumentation that betrayal of trust could bring about anger is also confirmed by Fehr et al. [76] with the data of an empirical experiment.

Third, emotion also effects on trust/distrust attitude. Dunn et al. [62]

report results from five experiments that describe the influence of emotional states on trust. They show that incidental emotions significantly influence trust in unrelated settings. Happiness and gratitude, emotions with positive valence, increase trust, and anger, an emotion with negative valence, decreases trust. Also, Andersen et al. [8] argue that positive emotions are a main factor in trust building, as they allow the actors to take the initial leap of faith, expecting that trust will be honored. McAllister [159] points out that trust based on emotional states such as care and concern is deeper than trust based primarily on predictability. A study of international strategic alliances by Cullen et al. [52] drew a distinction between credibility trust, which constitutes the rational component of trust building and benevolent trust, which constitutes the emotional dimension.

3.2 Trust in Computer Science

Trust is not a new research topic in computer science, spanning areas as diverse as security and access control in computer networks, reliability in distributed systems, game theory and agent systems, and policies for decision making under uncertainty. In this section, we aim to place the researches of trust into some main categories in order to provide a general view of researches of trust.

There are several attempts to classifier trust in computer science. For instance, Grandison and Sloman [93] have identified different forms of trust in the literature relating to whether access is being provided to the trustor's resources, the trustee is providing a service, trust concerns authentication. Ramchurn et al. [203], with respect to designing agents and open multi-agent systems, have conceptualized trust in two ways. The first is individual-level trust, whereby an agent has some beliefs about the honesty or reciprocative nature of its interaction partners. And the second way is system-level trust, whereby the actors in the system are forced to be trustworthy by the rules of encounter (i.e. protocols and mechanisms) that regulate the system. Artz and Gil [10] have distinguished four areas of trust: (i) Policy-based trust which uses policies to establish trust, focused on managing and exchanging credentials and enforcing access policies; (ii) Reputation-based trust which uses reputation to establish trust, where past interactions or performance for an entity are combined to assess its future behavior; (3) General models of trust which is a wealth of research on modeling and defining trust, its prerequisites, conditions, components, and consequences. Trust models are useful for analyzing human and agentized trust decisions and for operationalizing computable models of trust; and (4) trust in information resources which is an increasingly common theme in Web related research regarding whether Web resources and Web sites are reliable. Also, Sabater and Sierra [217] and Ries et al. [206] have listed computational models of trust based on the aspects of domain, dimension, and semantics of trust values. Ruohomaa and Kutvonen [214] took a survey on trust management. They divide research on trust into three groups based on its context, from low to high: infrastructure level, service level and community level. Mui et al. [170] have a review on the notation of reputation in multi-agent system. They consider reputation in the different scope: individual or group, and on the different mechanism of reputation: direct or indirect.

However, these classifications base on different points of view. Each of these classifications provide a narrow view on some particular aspects of trust: it is either trust management view, computational view, or based on the scope of trust or the mechanism trust is established on, etc. In order to give a general view on research of trust in computer science, we categorize research of trust on the ways it is built on. We thus distinguish three main categories of trust models:

- Protocol models: the models of trust oriented protocol, authentication, certification and security in the interactions between entities in the system or between computer and user.

- Computational models: the models concentrate on quantitative trust of a trustor in order to decide to trust or not on a given trustee.

- Cognitive models: the models concentrate on qualitative and formal representations of trust in order to recognize and reason about trust.

The next sections will present recent researches of trust in each category.

3.2.1 Protocol-oriented model of trust

This section presents the protocol-oriented models of trust, in which, trust is considered from the point of view of system management: authenticate, authority, security, etc. The objectives is to construct a model of trust which enables to establish security and trusted protocols. In these models, the trustee may be a user of a system, or a partner of an interaction.

Public key certificates A digital certificate is issued by a certification authority and verifies that a public key is owned by a particular entity. The certification authority does not vouch for the trustworthiness of the key owner, but simply authenticates the owner's identity. Pretty Good Privacy (PGP) trust model [174] is used for authentication relating to e-mail type of applications between human users. It supports a Web Of Trust model in that there is no centralized or hierarchical relationship between certification authorities as with X.509. The X.509 trust model [4, 241] is a strictly hierarchical trust model for authentication. Each entity must have a certificate that is signed by the central certification authority or another authority, which has been directly or indirectly certified by it. This model assumes that certification authorities are organized into a universal certification authority tree and that all certificates within a local community will be signed by a certification authority that can be linked into this tree (Blaze et al. [24]). The Platform for Content Selection (PICS) [166] was developed by the World Wide Web Consortium (W3C) as a solution to the problem of protecting children from pornography on the Internet without infringing on one's right to freedom of speech. PICS defines standards for the format and distribution of labels which are meta-document describing a Web document. Herzberg et al. [102] present a policy-based and certificate-based mechanism which can assign roles to new entrants. A certificate in this work is signed by some issuer and contains some claims about a subject. Mass et al. [157] extend the work in Herzberg et al. [102] to open multi-agent systems. Specifically, they take into account the fact that agents with reasoning or planning components can adapt their strategies rather than sticking to one strategy while maintaining their role. Poslad et al. [198] have recently proposed a number of security requirements they claim that are essential for agents to trust each other and each other's messages transmitted across the network linking them (i.e. to ensure messages are not tampered with by malicious agents). The authors specify these requirements for the FIPA (Foundation for Intelligent Physical Agents) abstract architecture [77].

Authenticate language Burrows et al. [33] propose a language to specify steps followed in the authentication process between two entities (resource access protocol analysis). The language is founded on cryptographic reasoning with logical operators defined to deal with notions of shared keys, public keys, encrypted statements, secrets, nonce freshness and statement jurisdiction (for authentication servers and certificate authorities). The Authorization Specification Language (ASL) proposed by Jajodia et al. [118] is used to specify authorization rules and makes explicit the need for the separation of policies and mechanisms. ASL supports the specification of the closed policy model (all allowable accesses must be specified) and the open policy model (all denied accesses must be explicitly specified) using a common architectural framework. It also supports role-based access control.

Trust management PolicyMaker is a trust management application, developed at AT&T Research Laboratories, that specifies what a public key is authorized to do [24]. KeyNote, the successor to PolicyMaker, was also developed by AT&T Research Laboratories to improve on the weaknesses of PolicyMaker. It has the same design principles of assertions and queries [23, 22] but includes two additional design goals, namely: standardization and ease of integration [22].

Additionally, the Rule-controlled Environment For Evaluation of Rules and Everything Else (REFEREE) is a trust management system for making access decisions relating to Web documents developed by Yang-Hua Chu based on PolicyMaker [45]. It considers a PICS label as the stereotypical web credential and uses the same theoretical framework as PolicyMaker to interpret trust policies and administer trust protocols, which are represented as software modules. Rangan [204] proposed a trust model for secure communication which is based on the logical secure channels between agents. The logical secure channels are established based on the mechanism of public keys schemes.

3.2.2 Computational model for trust

This section presents the computational models of trust, in which, trust is quantified or estimated. These models enable trustor to compare and to make a decision to trust on the trustee who has the highest trust value. They are popularly applied in the e-commerce application.

There are currently three main trends of researches in computational trust in multi agent system. Firstly, models of trust based on *personal experience* calculate the trust value of truster on trustee based on truster's personal parameters during transactions that have been performed with the trustee in the past. For instance, the model of Marsh [153] differentiates three types of trust: *Basic trust*, models the general trusting disposition independently of who is the agent that is in front. *General trust*, the trust that one agent has on another without taking into account any specific situation. *Situational trust*, the amount of trust that one agent has in another taking into account a specific situation. The probability theory-based model of Schillo et al. [223] is intended for scenarios where the result of an interaction between two agents is a boolean impression: *good* or *bad*; and there are no degrees of satisfaction. Manchala [149] proposes a model for the measurement of trust variables and the fuzzy verification of E-Commerce transactions, based on trust variables: cost of a transaction, transaction history, customer loyalty, indemnity and spending patterns. Nefti et al.'s [173] trust model considers the kind of information that is shown to increase customers trust when present on a merchant website. The model identified the following four major factors (modules): *existence*, *affiliation*, *policy* and *fulfillment*. After that, a fuzzy process will generate trust index for partners. The model of Abdul-Rahman and Hailes [1] uses four degrees of belief to typify agent trustworthiness: *vt (very trustworthy)*, *t (trustworthy)*, *u (untrustworthy)* and *vu (very untrustworthy)*. For each partner and context, the agent maintains a tuple with the number of past experiences in each category. Then, from the point of view of direct interaction, the trust on a partner in a given context is equal to the degree that corresponds to the maximum value in the tuple.

Secondly, models of trust based on reputation which estimate trust value based on truster's reference on her/his partners in their community or system. Their basic principle is that the more an agent is trusted by many other agents, the more trustworthy s/he is. For instance, online reputation models such as the model used on eBay [64], Amazon Auctions [7] and OnSale Exchange [185]. In the online reputation model Sporas [266], only the most recent rating between two users is considered. Another important characteristic is that users with very high reputation values experience much smaller rating changes after each update than users with a low reputation. This is a similar approach to the Glicko [88] system. Histos [266] was designed as a response to the lack of personalization that Sporas reputation values have. The model can deal with direct information (although in a very simple way) and witness information. Contrary to Sporas, the reputation value is a subjective property assigned particularly by each individual. In the model proposed by Yu and Singh [262, 263, 264], the information stored by an agent about direct interactions is a set of values that reflect the quality of these interactions. Only the most recent experiences with each concrete partner are considered for the calculations.

Thirdly, hybrid models makes use of integrating both personal experience and reference trust, and then final value of trust is aggregated by a weighted averaging operator. For instance, the model proposed by Esfandiari and Chandrasekharan [72] two one-on-one trust acquisition mechanisms are proposed. The first is based on observation and they propose the use of Bayesian networks to perform the trust acquisition by Bayesian learning. The second trust acquisition mechanism is based on interaction. The approach is the same one used in Lashkari et al. [134]. Sen and Sajja [228] reputation model, both types of direct experiences are considered: direct interaction and observed interaction. While the main idea behind the reputation model presented by Carter et al. [38] is that "the reputation of an agent is based on the degree of fulfillment of roles ascribed to it by the society". If the society judges that they have met their roles, they are rewarded with a positive reputation, otherwise they are punished with a negative reputation. Sabater and Sierra [215, 216] introduced ReGreT, a modular trust and reputation system oriented to complex small/mid-size e-commerce environments where social relations among individuals play an important role. The system takes into account three different sources of information: direct experiences, information from third party agents and social structures. The reputation model is divided in three specialized types of reputation depending on the information source that is used to calculate them: witness reputation, neighborhood reputation and system reputation. Ramchurn et al. [202] developed a trust model, based on confidence and reputation, and show how it can be concretely applied, using fuzzy sets, to guide agents in evaluating past interactions and in establishing new contracts with one another. They also use two kinds of trust: confidence, which is based on the direct interactions; and reputation, which is based on the mechanism of ReGReT system. Huynh et al. [113, 114] presented FIRE, a trust and reputation model that integrates a number of information sources to produce a comprehensive assessment of an agent's likely performance in open systems. Victor et al. [251] advocate the use of a trust model in which trust scores are (trust, distrust)-couples, drawn from a bilattice that preserves valuable trust provenance information including gradual trust, distrust, ignorance, and inconsistency. They focus on deriving trust information through reputation, in particular, the propagation of trust among agents. Tran and Nguyen [246] introduced a computational model of trust, which is also combination of experience and reference trust, based on fuzzy computational techniques and weighted aggregation operators. Teacy et al.'s model TRAVOS [244] calculates trust using probability theory taking account of past interactions between agents, and when there is a lack of personal experience between agents, the model draws upon reputation information gathered from third parties. Sensoy et al.'s model POYRAZ [229] combines a service selection engine that makes context-aware service selections using the shared consumer experiences and an information filtering module that computes trustworthiness of the consumers and identifies deceptive experiences. The model of Wang and Singh [254] understands trust in terms of belief and certainty. Certainty is formalized in terms of evidence based on a statistical measure defined over a probability distribution of the probability of positive outcomes.

Generally, computational models of trust enable agent to estimate concretely how much does s/he trust in a partner. Therefore, they are helpful for agents in situations where they need to make a decision to choose the most trustworthy partner to interact (collaboration, coordination, negotiation, etc.). However, in certain situations where agents need to reason to know why they trust a partner and/or what are possible consequences from an act of trust, how can agents prevent some negative consequence of trust from happening, the computational approaches of trust seem to be no more suitable: Agents need an approach which enables them to reason about trust. This requirement leads the cognitive formal approach of trust which will be presented in the next section.

3.2.3 Cognitive model for trust

This section presents the cognitive models of trust, in which, instead of being calculated, trust is qualitative represented based on belief, goal, etc. of trustor on trustee. In these models, trust is represented in the high-level formal primitives. They are applied to the applications of cognitive and reasoning systems.

Jones and Firozabadi. In [123], Jones and Firozabadi address the issue of the reliability of an agent's transmission. They use a modal logic of action developed by Kanger [127], Porn [196, 197] and Lindahl [143] to model agent's actions: $E_i p$ means "agent i brings it about that p". They use a variant of a normal modal logic of type KD45 (cf. Chellas [42]) as the foundation for their belief system: $B_i A$ means "agent i believes that A". Therefore, a deceitful communicative action is represented as follow:

$\neg B_a p \wedge E_a B_b E_a D \wedge B_a(E_a D \Rightarrow_b p)$

this says that a does not believe that p, that a sees to it that b believes that a sees to it that D, and that a believes that his seeing to it that D counts for b as sufficient evidence for the truth of p (cf. Jones and Firozabadi [123]). The formal representation brings out the idea that deceit is the exploitation of an assumed trust relation.

E2T2. Grandison and Reichgelt [92] introduce a model, called *Epistemic Event Temporal Trust* (E2T2), which enables the creation of trust-based relationships between online merchants and their consumers. They use an epistemic logic developed by von Wright [252] and Hintikka [108], and a temporal logic introduced by Venema [250] and Emerson [70]. Their formula of trust is follow:

$$
\begin{aligned}
&Trust(a, Calc(str_1, str_2, str_3, str_4), b, t, e) \leftrightarrow \\
&\qquad Bel(a, str_1, t, [(\forall x : T)(\forall x' : T)Bef(x, x')\wedge \\
&\qquad\qquad Bef(x', Tper(a, b, e, x)) \rightarrow \\
&\qquad\qquad Holds(CanInduce(b, e), x', Tper(a, b, e, x))]) \\
&\qquad \wedge Bel(a, str_2, t, (\forall x : T)[Occurs(Request(a, b, e), x) \rightarrow \\
&\qquad\qquad (\exists x' : T)(\exists x'' : T)[Bef(x, x')\wedge \\
&\qquad\qquad Bef(x', Tper(a, b, Request(a, b, e), x))\wedge
\end{aligned}
$$

$$
\begin{aligned}
&\quad Bef(x', x'') \wedge Occurs(Commit(b, e, Tper(a, b, e, x)), x'')]]) \\
&\wedge Bel(a, str_3, t, (\forall x : T)(\forall x' : T)[Bef(x, x') \wedge \\
&\quad Bef(x', Tper(a, b, e, x)) \rightarrow \\
&\quad Holds(Secure(b, e), x', Tper(a, b, e, x))]]) \\
&\wedge Bel(a, str_4, t, (\forall x : T)(\forall x' : T)[Occurs(Commit(b, e, x'), x) \rightarrow \\
&\quad (\exists x'' : T)[Bef(x, x') \wedge Bef(x', x'') \wedge Occurs(e, x'')]]])
\end{aligned}
$$

where $Trust(a, Calc(str_1, str_2, str_3, str_4), b, t, e)$ means that agent a trusts that agent b brings e at the time t with a tuple of four levels of belief in $Calc(str_1, str_2, str_3, str_4)$, which indicates that the trust level is dependent on the four quantified beliefs that make up the definition of trust: str_1, the level of a's belief about the competence of agent b in order to bring event e; str_2, the level of belief about the dependability, refers to reliability and timeliness, of agent b in order to bring event e; str_3, the level of agent a's belief that b will be secure with respect to e for the period in which it is reasonable for a to expect b to perform e after a has requested b to perform e; str_4, the level of belief that b is sincere with respect to its commitments. $Bel(a, str, t, \varphi)$ means that agent a believes φ is true at the time t with a level of str. $Bef(t_1, t_2)$ is true iff t_1 is earlier than t_2. $Holds(s, t_1, t_2)$ is true iff state s is true between time t_1 and t_2. $Occurs(e, t)$ is true iff event e occurs at time t. $Tper(a, b, e, x)$ means that agent b has committed to agent a to perform e at time t. $CanInduce(a, e)$ describes the state of a being able to bring about event e, either directly or through some third party. $Request(a, b, e)$ states that the first agent a requesting the second agent b to bring about the event e. $Commit(a, e, t)$ states that the agent a committing itself to bring about event e before that time t.

Herzig et al. Herzig et al. [103] provide a qualitative formal analysis of trust and reputation on the basis of cognitive primitives. The proposed formalization is strongly inspired by Falcone and Castelfranchi's model of social trust [75]. The concepts of trust and reputation are built from the same bricks (goal, capability, power, and willingness) but in a different scope (individual belief vs. collective belief).

Trust. They distinguish two types of trust: occurrent trust and dispositional trust. Occurrent trust is defined based on the components: belief, occurrent goal, occurrent capability, occurrent power and occurrent intention. The occurrent trust is formalized as follow:

$$\begin{aligned} OccTrust(i,j,\alpha,\varphi) \overset{def}{=} & OccGoal(i,\varphi) \wedge Bel_i OccCap(j,\alpha) \wedge \\ & Bel_i OccPower(j,\alpha,\varphi) \wedge Bel_i OccIntends(j,\alpha) \end{aligned}$$

where $OccTrust(i,j,\alpha,\varphi)$ means that agent i trust agent j in doing action α to achieve φ; $Bel_i\varphi$ reads "agent i believes that φ is true" based on the epistemic and doxastic logic (Hintikka [107]), the logic of belief is the modal system KD45 (cf. Chellas [42]); and:

- $OccGoal(i,\varphi) \overset{def}{=} Pref_i Eventually\varphi$ means that the occurrent goal of agent i is achievement of φ; where $Pref_i\varphi$ reads "agent i prefers φ is true", and $Eventually\varphi$ reads "eventually, φ is true" (cf. van Benthem [248]).

- $OccCap(j,\alpha) \overset{def}{=} \neg After_{j:\alpha}\bot$ represents the occurrent capacity of agent j in doing action α; where $After_{j:\alpha}\varphi$ expresses that φ is true after every possible execution of action α by agent j.

- $OccPower(j,\alpha,\varphi) \overset{def}{=} After_{j:\alpha}\varphi$ represents the power which relates j's action α with i's goal φ: j's performance of α will make φ true *here and now*.

- $OccIntends(j,\alpha) \overset{def}{=} Pref_j Does_{j:\alpha}\top$ represents the intention to do action α of agent j; where $Does_{j:\alpha}\varphi$ (introduced by Demolombe and Lorini [58]) means that agent j is performing action α.

The formula of dispositional trust is the same but the conditions are in the forms of potential conditions.

Reputation. They also use the same operators of goal, capacity, power and intention. But the group belief operator $GroupBelief_I\varphi$ is used instead of the individual belief operator.

$$
\begin{aligned}
Rep(I,j,\alpha,\varphi) \overset{def}{=} & OccGoal(I,\varphi) \wedge Bel_I OccCap(j,\alpha) \wedge \\
& Bel_I OccPower(j,\alpha,\varphi) \wedge Bel_I OccIntends(j,\alpha)
\end{aligned}
$$

where $Rep(I,j,\alpha,\varphi)$ reads "j has reputation in group I to do α in order to achieve φ"; $Bel_I\varphi$ reads "group I believes that φ is true".

This logic of trust is familial with our logic which will be presented in chapter 4. Thus we could basically inherit this approach to formalize the concept of trust in order to find out the effect of trust on emotions. The detail will be discussed in the section 5.2.2.

3.3 Conclusion

Trust has attracted many researches in recent years. There are various definitions, many models of trust have being proposed. In this chapter, we have categorized the models of trust in the literature based on the approach they are built on: Protocol models which model trust oriented protocol, authentication, certification and security in the interactions between entities in the system or between computer and user; computational models which concentrate on quantitative aspect of trust, these models enable trustor to make a decision to trust or not on a given trustee; and cognitive models which concentrate on qualitative aspect of trust, these models enable trustor to recognize and to reason about a given trustee's trustworthiness.

In oder to give the computer the ability to reason about the effect of trust and distrust on emotions, we concentrate on qualitative aspect of trust and distrust: we will use cognitive approach to construct a model of trust and distrust which will be presented in following chapters.

Chapter 4

Logical Framework

Three precedent chapters introduced the meaning of emotions, trust/distrust and their applications in computer science.

We now start to introduce our logical framework for emotions and trust/-distrust. This logic will enable us to lay bare the formal relations between trust and emotion, particularly the effect of trust and distrust on emotions. The logic we offer is a combination and extent of the logic of beliefs and choices of Herzig and Longin [105] (a refinement of the logic of choice and intention of Cohen and Levesque [50]), and the logic of time (introduced by Arthur Prior [201]).

Section 4.1 presents the syntax of the logic; section 4.2 presents the semantics of this logic; section 4.3 presents the axiomatics of logical operators in our logic and section 4.4 proves the decidability of our logic.

4.1 Syntax

4.1.1 Language

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... 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... denote such actions. The language of our logic is defined by the following BNF:

$$\begin{aligned}\varphi :=& p \mid i{:}\alpha\text{-}\mathit{happens} \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathtt{X}\varphi \mid \mathtt{X}^{-1}\varphi \mid \\ & \mathtt{G}\varphi \mid \mathtt{Bel}_i\,\varphi \mid \mathtt{Choice}_i\,\varphi \mid \mathtt{Grd}_I\varphi\end{aligned}$$

where p ranges over ATM, $i{:}\alpha$ ranges over $AGT \times EVT$, $i{:}\alpha\text{-}\mathit{happens}$ 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 $\bot$ (contradiction) are defined from $\neg$ (negation) and $\vee$ (disjunction).

- $i{:}\alpha\text{-}\mathit{happens}$ reads "agent i is just about to perform the action α";
- $\mathtt{X}\varphi$ reads "φ will be true next instant";
- $\mathtt{X}^{-1}\varphi$ reads "φ was true at the previous instant";
- $\mathtt{G}\varphi$ 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";
- $\mathtt{Grd}_I\varphi$ reads "φ is publicly grounded between the agents in group I".

4.1.2 Abbreviations

In this logic, we also define the following abbreviations:

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

$$\mathtt{Happens}_{i:\alpha}\varphi \overset{def}{=} i{:}\alpha\text{-}happens \wedge \mathtt{X}\varphi \qquad (\mathrm{Def}_{\mathtt{Happens}_{i:\alpha}})$$

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

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

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

$$\mathtt{Goal}_i\,\varphi \overset{def}{=} \mathtt{Choice}_i\,\mathtt{FBel}_i\,\varphi \qquad (\mathrm{Def}_{\mathtt{Goal}_i})$$

$$\mathtt{Intend}_i\,\alpha \overset{def}{=} \mathtt{Choice}_i\,\mathtt{F}i{:}\alpha\text{-}happens \qquad (\mathrm{Def}_{\mathtt{Intend}_i})$$

$$\mathtt{Able}_i\alpha \overset{def}{=} \neg\mathtt{After}_{i:\alpha}\bot \qquad (\mathrm{Def}_{\mathtt{Able}_i})$$

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

$$\mathtt{Aware}_i\varphi \overset{def}{=} \mathtt{X}^{-1}\neg\mathtt{Bel}_i\,\varphi \wedge \mathtt{Bel}_i\,\varphi \qquad (\mathrm{Def}_{\mathtt{Aware}_i})$$

where, the readability of these abbreviations is defined as follows:

- $i{:}\alpha\text{-}done$ reads "agent i has done action α";
- $\mathtt{Happens}_{i:\alpha}\varphi$ reads "agent i is doing action α and φ will be true next instant";
- $\mathtt{After}_{i:\alpha}\varphi$ reads "φ is true after any execution of α by i";
- $\mathtt{Done}_{i:\alpha}\varphi$ reads "agent i has done action α and φ was true at previous instant";
- $\mathtt{F}\varphi$ 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";
- $\mathtt{Intend}_i\,\alpha$ reads "agent i intends to do α";
- $\mathtt{Able}_i\alpha$ reads "agent i is capable to do α";
- $\mathtt{Poss}_i\varphi$ reads "agent i believes that it is possible φ";

- $\texttt{Aware}_i\varphi$ reads "agent i has just experienced that φ is true".

Sometimes, we also use some abbreviations revolving action α to represent the cases of *inaction*:

$$\texttt{Done}_{\sim i:\alpha}\varphi \overset{def}{=} \neg i{:}\alpha\text{-}done \wedge \texttt{X}^{-1}\varphi \tag{$\mathrm{Def}_{\texttt{Done}_{\sim i:\alpha}}$}$$

$\texttt{Done}_{\sim i:\alpha}\varphi$ reads "agent i did not do action α and φ was true at previous instant".

4.2 Semantics

For temporal operators, we use a semantics based on linear time described by a story (or sequence) of time points (This semantics is very close to CTL*, cf. Clarke et al. [47]).

4.2.1 Frame

Let W set of possible worlds. A frame $\mathcal{F}$ is a 4-tuples $\langle \mathscr{H}, \mathscr{B}, \mathscr{C}, \mathscr{G} \rangle$ where:

- $\mathscr{H}$ is a set of situations 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 (h, z)). Each situation (h, z) is correspond to a possible world $w \in W$.

- $\mathscr{B}$ is the set of all $\mathscr{B}_i$ such that $\mathscr{B}_i(h, z) \subseteq \mathscr{H}$ denotes the set of stories believed as being possible by the agent i in the situation (h, z). All the accessibility relations $\mathscr{B}_i$ are serial[1], transitive[2] and Euclidean[3]. This semantic is completely standard in epistemic logic (cf. Gochet and Gribomont [89] and Hintikka [107]).

- $\mathscr{C}$ is the set of all $\mathscr{C}_i$ such that $\mathscr{C}_i(h, z) \subseteq \mathscr{H}$ denotes the set of stories chosen by the agent i in the situation (h, z). All the accessibility $\mathscr{C}_i$ are serial.

[1] for every $i \in AGT$, $\mathscr{B}_i(h, z) \neq \emptyset$
[2] if $h_1 \in \mathscr{B}_i(h, z)$ and $h_2 \in \mathscr{B}_i(h_1, z)$, then $h_2 \in \mathscr{B}_i(h, z)$
[3] if $h_1 \in \mathscr{B}_i(h, z)$ and $h_2 \in \mathscr{B}_i(h, z)$, then $h_2 \in \mathscr{B}_i(h_1, z)$

- $\mathscr{G}$ is the set of all $\mathscr{G}_I$ such that $\mathscr{G}_I(h,z) \subseteq \mathscr{H}$ denotes the set of stories which are publicly grounded in the group I of agents, in the situation (h,z). All the accessibility relations $\mathscr{G}_I$ are serial, transitive and Euclidean (This is similar to the operator group grounding introduced by Gaudou et al. [85]).

Moreover, we impose for every $z \in \mathbb{Z}$ that: if $h_1 \in \mathscr{B}_i(h,z)$ then $\mathscr{C}_i(h,z) = \mathscr{C}_i(h_1,z)$. It means that if an agent believes that the situation (h_1,z) is possible from the situation (h,z), then the set of his/her preference worlds from (h,z) and (h_1,z) 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).

4.2.2 Model & Truth conditions

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 $(h,z) \in \mathcal{V}(p)$

$\mathcal{M}, h, z \models \neg\varphi$ iff $\mathcal{M}, h, z \not\models \varphi$

$\mathcal{M}, h, z \models \varphi \vee \psi$ iff $\mathcal{M}, h, z \models \varphi$ or $\mathcal{M}, h, z \models \psi$

$\mathcal{M}, h, z \models \mathtt{X}\varphi$ iff $\mathcal{M}, h, z+1 \models \varphi$

$\mathcal{M}, h, z \models \mathtt{X}^{-1}\varphi$ iff $\mathcal{M}, h, z-1 \models \varphi$

$\mathcal{M}, h, z \models \mathtt{G}\varphi$ iff $\mathcal{M}, h, z' \models \varphi$ for every $z' \geq z$

$\mathcal{M}, h, z \models \mathtt{Bel}_i\,\varphi$ iff $\mathcal{M}, h', z \models \varphi$ for every $h' \in \mathscr{B}_i(h,z)$

$\mathcal{M}, h, z \models \mathtt{Choice}_i\,\varphi$ iff $\mathcal{M}, h', z \models \varphi$ for every $h' \in \mathscr{C}_i(h,z)$

$\mathcal{M}, h, z \models \mathtt{Grd}_I\varphi$ iff $\mathcal{M}, h', z \models \varphi$ for every $h' \in \mathscr{G}_I(h,z)$

4.3 Axiomatics

4.3.1 Temporal operators

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

$$i{:}\alpha\text{-}happens \leftrightarrow \mathtt{X}\, i{:}\alpha\text{-}done \tag{4.1}$$

$$\mathtt{X}\varphi \leftrightarrow \neg\mathtt{X}\neg\varphi \tag{4.2}$$

$$\varphi \leftrightarrow \mathtt{X}\mathtt{X}^{-1}\varphi \tag{4.3}$$

$$\varphi \leftrightarrow \mathtt{X}^{-1}\mathtt{X}\varphi \tag{4.4}$$

$$\mathtt{G}\varphi \leftrightarrow \varphi \wedge \mathtt{X}\mathtt{G}\varphi \tag{4.5}$$

$$\mathtt{G}(\varphi \to \mathtt{X}\varphi) \to (\varphi \to \mathtt{G}\varphi) \tag{4.6}$$

Axiom (4.1) is straightforward from the definition of the abbreviation ($\mathrm{Def}_{i:\alpha\text{-}done}$); axiom (4.2) means that the next instant operator X is not contradictory: it is not the case that φ and $\neg\varphi$ are both true in the next instant; axioms (4.3) and (4.4) mean that the next instant operator X is temporally symmetrical; axioms (4.5) and (4.6) represent the recursive definition of $G\varphi$.

4.3.2 Actions

Linear time semantics also entail the following principles:

$$\mathtt{G}\varphi \to \mathtt{After}_{i:\alpha}\varphi \tag{4.7}$$

$$\mathtt{Happens}_{i:\alpha}\varphi \to \mathtt{After}_{j:\beta}\varphi \tag{4.8}$$

$$\mathtt{After}_{i:\alpha}\varphi \leftrightarrow \neg\mathtt{Happens}_{i:\alpha}\neg\varphi \tag{4.9}$$

Axiom (4.7) 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 (h, z), and that φ be false at time $z' > z$.)

As time is linear, actions are deterministic on a given history. Thus, axiom (4.8) 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.

Axiom (4.9) means that $\texttt{After}_{i:\alpha}$ and $\texttt{Happens}_{i:\alpha}$ operators are dual operators. This property is fair with respect to dynamic logic (Harel et al. [100]).

4.3.3 Belief

$\texttt{Bel}_i$ operator is defined in a logic KD45 (cf. Chellas [42]):

$$\frac{\varphi}{\texttt{Bel}_i\,\varphi} \tag{$\mathrm{RN}_{\texttt{Bel}_i}$}$$

$$\texttt{Bel}_i\,(\varphi \to \psi) \to (\texttt{Bel}_i\,\varphi \to \texttt{Bel}_i\,\psi) \tag{$\mathrm{K}_{\texttt{Bel}_i}$}$$

$$\texttt{Bel}_i\,\varphi \to \neg\texttt{Bel}_i\,\neg\varphi \tag{$\mathrm{D}_{\texttt{Bel}_i}$}$$

$$\texttt{Bel}_i\,\varphi \leftrightarrow \texttt{Bel}_i\,\texttt{Bel}_i\,\varphi \tag{$4_{\texttt{Bel}_i}$}$$

$$\neg\texttt{Bel}_i\,\varphi \leftrightarrow \texttt{Bel}_i\,\neg\texttt{Bel}_i\,\varphi \tag{$5_{\texttt{Bel}_i}$}$$

Axiom ($\mathrm{RN}_{\texttt{Bel}_i}$) means that all theorems are believed by every agent i; axiom ($\mathrm{K}_{\texttt{Bel}_i}$) means that beliefs are closed under material implication for every agent i; axiom ($\mathrm{D}_{\texttt{Bel}_i}$) means that beliefs of every agent i are rational: they cannot be contradictory; axioms ($4_{\texttt{Bel}_i}$) and ($5_{\texttt{Bel}_i}$) mean that agent i is conscious of its beliefs and of its disbeliefs.

4.3.4 Choice

$\texttt{Choice}_i$ operators is also defined in a normal modal logic plus (D) axioms (cf. Chellas [42]):

$$\frac{\varphi}{\mathtt{Choice}_i\,\varphi} \tag{$\mathrm{RN}_{\mathtt{Choice}_i}$}$$

$$\mathtt{Choice}_i\,(\varphi \rightarrow \psi) \rightarrow (\mathtt{Choice}_i\,\varphi \rightarrow \mathtt{Choice}_i\,\psi) \tag{$\mathrm{K}_{\mathtt{Choice}_i}$}$$

$$\mathtt{Choice}_i\,\varphi \rightarrow \neg\mathtt{Choice}_i\,\neg\varphi \tag{$\mathrm{D}_{\mathtt{Choice}_i}$}$$

Axiom ($\mathrm{RN}_{\mathtt{Choice}_i}$) means that all theorems are chosen by every agent i; axiom ($\mathrm{K}_{\mathtt{Choice}_i}$) means that choices are closed under material implication for every agent i; axiom ($\mathrm{D}_{\mathtt{Choice}_i}$) means that choices of every agent i are rational: they cannot be contradictory.

The following principle follows from the semantical constraint between belief accessibility relation and choice accessibility relation, and from axiom ($\mathrm{D}_{\mathtt{Bel}_i}$):

$$\mathtt{Choice}_i\,\varphi \leftrightarrow \mathtt{Bel}_i\,\mathtt{Choice}_i\,\varphi \tag{4_{BC}}$$

$$\neg\mathtt{Choice}_i\,\varphi \leftrightarrow \mathtt{Bel}_i\,\neg\mathtt{Choice}_i\,\varphi \tag{5_{BC}}$$

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

4.3.5 Public ground

The sound and complete axiomatization of $\mathtt{Grd}_I$ 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 [253], also introduced by Gaudou et al. [85]:

$$\frac{\varphi}{\mathtt{Grd}_I\varphi} \tag{$\mathrm{RN}_{\mathtt{Grd}_I}$}$$

$$\mathtt{Grd}_I(\varphi \rightarrow \psi) \rightarrow (\mathtt{Grd}_I\varphi \rightarrow \mathtt{Grd}_I\psi) \tag{$\mathrm{K}_{\mathtt{Grd}_I}$}$$

$$\mathtt{Grd}_I\varphi \rightarrow \neg\mathtt{Grd}_I\neg\varphi \tag{$\mathrm{D}_{\mathtt{Grd}_I}$}$$

$$\mathtt{Grd}_I\varphi \rightarrow \mathtt{Grd}_I\mathtt{Grd}_I\varphi \tag{$4_{\mathtt{Grd}_I}$}$$

$$\neg\mathtt{Grd}_I\varphi \rightarrow \mathtt{Grd}_I\neg\mathtt{Grd}_I\varphi \tag{$5_{\mathtt{Grd}_I}$}$$

Axiom ($\mathrm{RN}_{\mathtt{Grd}_I}$) means that every tautology is public ground. Axiom ($\mathrm{K}_{\mathtt{Grd}_I}$) means that if φ is publicly grounded in I and that φ implies ψ then ψ is also publicly grounded in I. Axiom ($\mathrm{D}_{\mathtt{Grd}_I}$) 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 ($4_{\mathtt{Grd}_I}$) and negative introspection axiom ($5_{\mathtt{Grd}_I}$) account for the public character of $\mathtt{Grd}_I$. 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.

4.4 Soundness and completeness

We call $\mathcal{L}$ the logic axiomatized, and write $\vdash_{\mathcal{L}} \varphi$ iff φ is a theorem of $\mathcal{L}$.

Theorem 1 (Soundness and completeness). $\vdash \varphi$ *iff* $\models \varphi$

We here present the proof of axiomatic system KD45 for *Belief* and *Ground*, KD45-like axiomatic system for *Choice*. The axiomatic system for temporal operators could refer from the work of Clarke et al. [47]. The axiomatic system for dynamic operators could refer from the work of Harel et al. [100].

4.4.1 Soundness and Completeness for *Belief*

Soundness of system KD45 for Belief

Soundness of axiom K of Belief. We need to indicate that if $\mathtt{Bel}_i\,\varphi$ and $\mathtt{Bel}_i\,(\varphi \rightarrow \psi)$ hold in (h, z) then $\mathtt{Bel}_i\,\psi$ will also hold in (h, z).

Suppose that $\mathtt{Bel}_i\,\varphi$ and $\mathtt{Bel}_i\,(\varphi \rightarrow \psi)$ hold in (h, z), so from the truth condition of belief operator, φ and $(\varphi \rightarrow \psi)$ hold in all situations (h_1, z) such that $h_1 \in \mathscr{B}_i(h, z)$. From material implication relation, we have that ψ holds in all situations (h_1, z) such that $h_1 \in \mathscr{B}_i(h, z)$. Therefore $\mathtt{Bel}_i\,\psi$ holds in (h, z) (from the truth condition of $\mathtt{Bel}_i\,\psi$).

$\square$

Soundness of axiom D of Belief. We need to indicate that if $\texttt{Bel}_i\,\varphi$ holds in (h, z) then $\neg\texttt{Bel}_i\,\neg\varphi$ will also hold in (h, z).

Suppose that $\texttt{Bel}_i\,\varphi$ holds in (h, z), so from the truth condition of belief operator, φ holds in all situations (h_1, z) such that $h_1 \in \mathscr{B}_i(h, z)$ (i).

Contradictory suppose that $\neg\texttt{Bel}_i\,\neg\varphi$ does not hold in (h, z), it means that $\texttt{Bel}_i\,\neg\varphi$ holds in (h, z). So, from the truth condition of belief operator, $\neg\varphi$ holds in all situations (h_1, z) such that $h_1 \in \mathscr{B}_i(h, z)$ (ii).

Recall that $\mathscr{B}_i$ is serial, therefore $\mathscr{B}_i(h, z) \neq \emptyset$. So, from (i) and (ii) there is a conflict in all situation (h_1, z) such that $h_1 \in \mathscr{B}_i(h, z)$. Thus, the contradictory hypotheses is wrong! That is all we need to show!

□

Soundness of axiom 4 of Belief. By contrapossition of ($4_{\texttt{Bel}_i}$), we just need to show that if $\neg\texttt{Bel}_i\,\texttt{Bel}_i\,\varphi$ holds in (h, z) then $\neg\texttt{Bel}_i\,\varphi$ will also hold in (h, z).

Suppose that $\neg\texttt{Bel}_i\,\texttt{Bel}_i\,\varphi$ holds in (h, z), so from the truth condition of belief operator, $\texttt{Bel}_i\,\varphi$ does not hold in at least a situation (h_1, z) such that $h_1 \in \mathscr{B}_i(h, z)$. Thus, φ does not hold in at least a situation (h_2, z) such that $h_2 \in \mathscr{B}_i(h_1, z)$ by the truth condition of $\texttt{Bel}_i\,\varphi$. But $\mathscr{B}_i$ is transitive so from $h_1 \in \mathscr{B}_i(h, z)$ and $h_2 \in \mathscr{B}_i(h_1, z)$, we have $h_2 \in \mathscr{B}_i(h, z)$. Thus, we can say φ does not hold in at least a situation (h_2, z) such that $h_2 \in \mathscr{B}_i(h, z)$. It means $\texttt{Bel}_i\,\varphi$ does not hold in situation (h, z), so $\neg\texttt{Bel}_i\,\varphi$ holds.

□

Soundness of axiom 5 of Belief. By contrapossition of ($5_{\texttt{Bel}_i}$), we just need to show that if $\neg\texttt{Bel}_i\,\neg\texttt{Bel}_i\,\varphi$ holds in (h, z) then $\texttt{Bel}_i\,\varphi$ will also hold in (h, z).

Suppose that $\neg\texttt{Bel}_i\,\neg\texttt{Bel}_i\,\varphi$ holds in (h, z), thus $\texttt{Bel}_i\,\varphi$ holds in at least a situation (h_1, z) such that $h_1 \in \mathscr{B}_i(h, z)$. This means that φ holds in all situations (h_2, z) such that $h_2 \in \mathscr{B}_i(h_1, z)$ (iii).

Contradictory suppose that $\texttt{Bel}_i\,\varphi$ does not hold in (h, z). So from the truth condition of belief operator, φ does not hold in at least a situation (h_3, z) such that $h_3 \in \mathscr{B}_i(h, z)$. But $\mathscr{B}_i$ is Euclidean so from $h_3 \in \mathscr{B}_i(h, z)$

and $h_1 \in \mathscr{B}_i(h, z)$, we have $h_3 \in \mathscr{B}_i(h_1, z)$. Therefore, we can say φ does not hold in at least a situation (h_3, z) such that $h_3 \in \mathscr{B}_i(h_1, z)$ (iv).

(iii) conflicts with (iv), so the contradictory hypotheses is wrong, thus $\texttt{Bel}_i\, \varphi$ holds in (h, z). That all we need to show.

□

Completeness of system KD45 for Belief

Completeness of system KD45 for Belief. We need to show that system K-D45 is complete with respect to the class of all transitive and Euclidean frames.

Given a KD45-consistent set of formulas $\sum$, we will show the contrapositive: let ϕ any formulas in $\sum$, we will show: if $\sum \nvdash \phi$ then $\sum \not\models \phi$.

From $\sum \nvdash \phi$, this implies $\sum \cup \neg\phi$ is consistent. By the Lindenbaum's Lemma (cf. Blackburn et al. [21]) there exist a maximal consistent set $\sum^+$ extending $\sum \cup \neg\phi$. By the Existence Lemma (cf. Blackburn et al. [21]) there exist a model in $\sum^+$ extending $\sum \cup \neg\phi$. By the Truth Lemma (cf. Blackburn et al. [21]) we have $\sum \cup \neg\phi \models \neg\phi$, so $\sum \cup \neg\phi \not\models \phi$.

For the class of all transitive frames. Consider the canonical model $(\mathscr{H}, \mathcal{R}, \mathcal{V})$ for KD45. Suppose $(h_1, z), (h_2, z), (h_3, z)$ are situations in $(\mathscr{H}, \mathcal{R}, \mathcal{V})$ such that $h_1 \in \mathcal{R}(h_2, z)$ and $h_2 \in \mathcal{R}(h_3, z)$. We need to show that $h_1 \in \mathcal{R}(h_3, z)$.

Suppose that φ holds in (h_3, z), from $h_2 \in \mathcal{R}(h_3, z)$ then it must be that $\Diamond\varphi$ holds in (h_2, z). And from $h_1 \in \mathcal{R}(h_2, z)$ then it must be that $\Diamond\Diamond\varphi$ holds in (h_1, z). But (h_1, z) is a maximal consistent set, so $\Diamond\Diamond\varphi \to \Diamond\varphi$ holds in (h_1, z). By modus ponens, $\Diamond\varphi$ also holds in (h_1, z), so $h_1 \in \mathcal{R}(h_3, z)$.

For the class of all Euclidean frames. Suppose $(h_1, z), (h_2, z), (h_3, z)$, and (h_4, z) are situations in $(\mathscr{H}, \mathcal{R}, \mathcal{V})$ such that $h_2 \in \mathcal{R}(h_1, z)$, $h_3 \in \mathcal{R}(h_1, z)$ and $h_4 \in \mathcal{R}(h_3, z)$. We will show that $h_4 \in \mathcal{R}(h_2, z)$ and then $h_3 \in \mathcal{R}(h_2, z)$ or/and $h_2 \in \mathcal{R}(h_3, z)$.

Suppose that φ holds in (h_1, z), from $h_2 \in \mathcal{R}(h_1, z)$ then it must be that $\Diamond\varphi$ holds in (h_2, z). And from $h_3 \in \mathcal{R}(h_1, z)$ then it must be that $\Diamond\varphi$ also holds in (h_3, z).

Otherwise, from $h_4 \in \mathcal{R}(h_3, z)$ and $\Diamond\varphi$ holds in (h_3, z) then $\Diamond\Diamond\varphi$ holds in (h_4, z).

From $\Diamond\varphi$ holds in (h_2, z) and $\Diamond\Diamond\varphi$ holds in (h_4, z), then we also have relation $h_4 \in \mathcal{R}(h_2, z)$.

But (h_4, z) is a maximal consistent set, so $\Diamond\Diamond\varphi \to \Diamond\varphi$ holds in (h_4, z). By modus ponens, $\Diamond\varphi$ also holds in (h_4, z) therefore φ also holds in (h_2, z) and (h_3, z).

From $\Diamond\varphi$ holds in (h_2, z) and φ holds in (h_3, z), then we also have relation $h_2 \in \mathcal{R}(h_3, z)$. And also, from $\Diamond\varphi$ holds in (h_3, z) and φ holds in (h_2, z), then we also have relation $h_3 \in \mathcal{R}(h_2, z)$.

□

4.4.2 Soundness and Completeness for *Ground*

The soundness and completeness of axiomatic system KD45 for *Ground* could be proved in the same principle to the *Belief* operator (cf. section 4.4.1), by replacing $\mathscr{B}_i(h, z)$ by $\mathscr{G}_I(h, z)$ where $h = h_1, h_2, h_3$ respectively.

4.4.3 Soundness and Completeness for *Choice*

The soundness of axiomatic system KD and completeness of axiomatic system KD45-like for *Choice* could be proved in the same principle to the *Belief* operator (cf. section 4.4.1), by replacing $\mathscr{B}_i(h, z)$ by $\mathscr{C}_i(h, z)$ where $h = h_1, h_2, h_3$ respectively. We here present the proof of axiomatic 4_{BC} and 5_{BC} for *Choice*.

Soundness of axiom 4_{BC} of Choice. By contrapossition of (4_{BC}), we just need to show that if $\neg\texttt{Bel}_i\,\texttt{Choice}_i\,\varphi$ holds in (h, z) then $\neg\texttt{Choice}_i\,\varphi$ will also hold in (h, z).

Suppose that $\neg \mathtt{Bel}_i\, \mathtt{Choice}_i\, \varphi$ holds in (h, z), so from the truth condition of belief operator, $\mathtt{Choice}_i\, \varphi$ does not hold in at least a situation (h_1, z) such that $h_1 \in \mathscr{B}_i(h, z)$. Thus φ does not hold in at least a situation (h_2, z) such that $h_2 \in \mathscr{C}_i(h_1, z)$ by the truth condition of choice operator.

Note that we assume that if $h_1 \in \mathscr{B}_i(h, z)$, then $\mathscr{C}_i(h_1, z) = \mathscr{C}_i(h, z)$ (cf. section 4.2.1). So from $h_2 \in \mathscr{C}_i(h_1, z)$ and $\mathscr{C}_i(h_1, z) = \mathscr{C}_i(h, z)$, we have $h_2 \in \mathscr{C}_i(h, z)$.

Therefore, we can say φ does not hold in at least a situation (h_2, z) such that $h_2 \in \mathscr{C}_i(h, z)$. It means $\mathtt{Choice}_i\, \varphi$ does not hold in situation (h, z), so $\neg \mathtt{Choice}_i\, \varphi$ holds.

□

Soundness of axiom 5_{BC} of Choice. By contrapossition of (5_{BC}), we just need to show that if $\neg \mathtt{Bel}_i\, \neg \mathtt{Choice}_i\, \varphi$ holds in (h, z) then $\mathtt{Choice}_i\, \varphi$ will also hold in (h, z).

Suppose that $\neg \mathtt{Bel}_i\, \neg \mathtt{Choice}_i\, \varphi$ holds in (h, z), thus $\mathtt{Choice}_i\, \varphi$ holds in at least a situation (h_1, z) such that $h_1 \in \mathscr{B}_i(h, z)$. This means that φ holds in all situations (h_2, z) such that $h_2 \in \mathscr{C}_i(h_1, z)$ by the truth condition of choice operator (v).

Contradictory suppose that $\mathtt{Choice}_i\, \varphi$ does not hold in (h, z). So from the truth condition of choice operator, φ does not hold in at least a situation (h_3, z) such that $h_3 \in \mathscr{C}_i(h, z)$. But note that we assume that if $h_1 \in \mathscr{B}_i(h, z)$, then $\mathscr{C}_i(h_1, z) = \mathscr{C}_i(h, z)$ (cf. section 4.2.1). So from $h_3 \in \mathscr{C}_i(h, z)$ and $\mathscr{C}_i(h_1, z) = \mathscr{C}_i(h, z)$, we have $h_3 \in \mathscr{C}_i(h_1, z)$. Therefore, we can say φ does not hold in at least a situation (h_3, z) such that $h_3 \in \mathscr{C}_i(h_1, z)$ (vi).

Recall that $\mathscr{C}_i$ is serial, then $\mathscr{C}_i \neq \emptyset$. Thus, (v) conflicts with (vi), so the contradictory hypotheses is wrong, thus $\mathtt{Choice}_i\, \varphi$ holds in (h, z). That all we need to show.

□

4.5 Complexity

The complexity of our logic could be determined via those of its operators' axiomatic systems.

Temporal operators $\mathtt{X}\varphi, \mathtt{X}^{-1}\varphi$, and $\mathtt{G}\varphi$. Their semantic is linear time likely the semantic of CTL* (cf. Clarke et al. [47]) whose complexity is well known as PSPACE for branching semantic (cf. Emerson et al. [71] and Schnoebelen [224]). Because our semantic is linear time instead of branching time, so it is clearly that their complexity is not more complex than polynomial-time.

$\mathtt{Bel}_i$ operator. Its semantic is the logic KD45 whose complexity is also well known like PSPACE (cf. Halpern and Moses [98], Massacci [158] and Shvarts [232]).

$\mathtt{Choice}_i$ operator. Its semantic is a logic KD45-like (cf. Herzig and Longin [105], Cohen and Levesque [50]), so we could consider its complexity as those of logic KD45 whose complexity is also well known like PSPACE (cf. Halpern and Moses [98], Massacci [158] and Shvarts [232]).

$\mathtt{Grd}_i$ operator. Its semantic is also a logic KD45-like (cf. Gaudou et al. [85]), so we could consider its complexity as those of logic KD45 whose complexity is also well known like PSPACE (cf. Halpern and Moses [98], Massacci [158] and Shvarts [232]).

4.6 Conclusion

This chapter introduced a propositional modal logic to represent the agents' mental attitudes and reasoning abilities. Therefore, agent has beliefs, personal choices and public ground with others, and s/he can reason about time and action. The introduced logic is an extent of the logical framework of Herzig and Longin [105] (a refinement of the logic of choice and intention of Cohen and Levesque [50]) with both *public ground* and *next* operators. The semantic of action and time is also quite different.

This logic will be used to formalize the concept of emotion and trust/distrust (chapter 5) as well as some relation between them, particularly the effects of trust/distrust on emotions (chapter 6).

Chapter 5

Formalization of Emotion and Trust/Distrust

This chapter uses the logic introduced in chapter 4 to formalize the concept of some emotions and trust/distrust. Our aims is not to formalize these concepts in general context but in some certain context, based on the point of view of cognitivists who argue that each emotion (resp. trust/distrust) has some particular cognitive factors such that their combination enables to recognize and distinguish from other emotions. Thus, our formalization is really the formalization of the cognitive structure of emotion (resp. trust/distrust).

Section 5.1 presents the formalization of the cognitive structure of some emotions; Section 5.2 presents the formalization of the cognitive structure of trust and distrust.

5.1 Formalization of Emotions

5.1.1 Choice of foundation models

In this section, we use our logic to formalize the cognitive structure of some emotions. In particular, we mainly base on *the cognitive structure* as proposed by Ortony et al. [186], *the situational meaning structures* as called by Frijda [81] and *the appraisal patterns* proposed by Lazarus [135]. We also refer to *the belief-desire theory of emotion* of Reisenzein [205] as well as the model of Scherer et al. [222]. There are three main reasons why we chosen these theories of emotion as foundation models on which we construct our logical model of emotions.

Firstly, each of them is a cognitive model, in which each emotion is considered as a unique combination of its decompositions (cognitive factors). This combination is called *the cognitive structure* as proposed by Ortony et al. [186], or *the situational meaning structures* as called by Frijda [81], or *the appraisal patterns* proposed by Lazarus [135]. Hereafter, we call it as the cognitive structure of each emotion. This cognitive structure enables to recognize, to distinguish and to reason about each particular emotion. This is useful for us to formalize the concept of each emotion because our objective is also make up to the computer the ability to recognize, to distinguish and to reason about each particular emotion.

Secondly, these theories support AI applications. As Ortony et al. mentioned in their book "we would like to lay the foundation for a computationally tractable model of emotion. In other words, we would like an account of emotion that could in principle be used in an Artificial Intelligence system that would, for example, be able to reason about emotion."[186, p. 2]. Moreover, these theories propose some common cognitive factor of emotion such as *desirability* (cf. Ortony et al. [186]) *goal* (cf. Lazarus [135]) which are closed to some atomic operator in modal logic such as *desire* in BDI logic of Georgeff [86] or *goal* in Cohen and Levesque's logic [50]. Thus, it seem to be helpful to us to construct our atomic operators from these cognitive factors because our logic is also based on the logic of BDI [86] and the logic of Cohen and Levesque [50].

Thirdly, they are popular applied in the computer sciences, particularly in AI applications. Many emotional models for computer are based on these theories. For instance, the model of Adam et al. [2, 3], Ochs et al. [181, 182], Steunebrink et al. [238, 239], Gershenson [87], Van Dyke Parunak et al. [191] and the model of Yao et al. [261]. They are all based on the theory of Ortony et al. [186].

The following sections will present the formalization of emotions based on these foundation models. We also divide emotions into some categories corresponding to those of Ortony et al. [186]: well-being emotions, prospect-based emotions, confirmation-based emotion and compound emotions.

5.1.2 Well-being emotions: Joy/Distress

Ortony et al. [186] argue that we feel joy when we are pleased about a desirable event. Lazarus [135] argues that when we have a goal congruence, and the overall life outlook is favorable, then that is essential to feel joy. Frijda [81] proposed that the situational meaning structure of joy is either one of accomplished striving or one in which the way to accomplishment appears open. Reisenzein [205] argues that when what we desire is certain, we then feel joy. We accordingly summarize that 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[1] formalized as $\mathtt{Goal}_i\,\varphi$, and the second factor can be formalized as $\mathtt{Bel}_i\,\varphi$.

[1]We put *"potentially"* because it could be done like this but we then analyze more detailed in order to obtain a more precised formula

However, considering the time factor, we need a more detailed analysis. In particular, the first factor of *Joy* means that agent i now recalls that in the previous instant, he desired φ (until experiencing that φ was in fact true) which can be written: $\texttt{Bel}_i\,\texttt{X}^{-1}\texttt{Goal}_i\,\varphi$. In other term, in order to be joyful, agent i must keep in mind his desire in the previous instant. It is clear that if agent forgot (or changed his chosen preference), he would not feel joy when the associated proposition becomes true. Hereafter, we add this analysis for almost emotional formulas.

For the second factor of *Joy*, we assume that emotion is triggered at the moment when all its (cognitive) factors are fulfilled, and that its intensity then decreases with time as it has been argued by de Sousa [56], and Frijda [81].

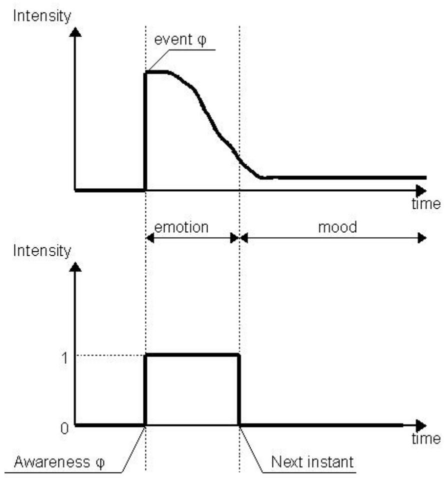


Figure 5.1: Formalization of emotion based on time factor.

Figure 5.1 illustrates the formalization of emotions based on time. In the chart on top, the natural intensity of emotion is maximal when the associated event occurs. It is then decreased following the time and finally, it becomes the mood (cf. section 1.5) of the same emotion [56, 81]. For instance, when you apply for a job: you feel joyful when your new boss informs to you that you are recruited, that rightly your emotion! But one year latter, do you always feel joy? It is not right: Your feeling of joy at that time, if it exists, is only the mood status of the emotion having the same name!

The bottom chart in Figure 5.1 illustrates our formalization of corresponding emotion: the emotion is present only in the current time slot, from the awareness of associated event until the next instant. Emotion does not exist forever.

Accordingly, we include this time factor into most emotional formulas. Thus, the second factor of *Joy* means that agent i has just experienced that φ is true and did not previously know it, that is: $\texttt{Aware}_i\varphi$ (cf. section 4.1). This interpretation enables to eliminate the case where the agent i's joy exists until the next instant because in the next instant, agent i might always know that φ is true ($\texttt{Bel}_i\,\varphi$) but he has already known it since previous instant. So it does not means that i has experienced that φ is true ($\neg\texttt{Aware}_i\varphi$). Thus the second factor is not fulfilled, and his joy does not exist anymore!

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

Definition 1. *(Joy/Distress)*

$$\texttt{Joy}_i\,\varphi \overset{def}{=} \texttt{Bel}_i\,\texttt{X}^{-1}\texttt{Goal}_i\,\varphi \wedge \texttt{Aware}_i\varphi$$
$$\texttt{Distress}_i\,\varphi \overset{def}{=} \texttt{Bel}_i\,\texttt{X}^{-1}\texttt{Goal}_i\,\neg\varphi \wedge \texttt{Aware}_i\varphi$$

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

5.1.3 Prospect-based emotions: Hope/Fear

Ortony et al. [186] argue that we are hopeful when we are pleased about the prospect of a desirable event. Lazarus [135] argues that when we have a goal congruence, and the future expectations must be uncertain, then we are hopeful. Searle [227] argues that hope is composed of one's goal (or wish) and possibility of related event. Reisenzein [205] argues that when what we desire is uncertain, we then hope about it. Miceli and Castelfranchi [165] argue that we hope about φ when we have a goal φ, and φ is possible and outside of our control. We accordingly summarize that 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 agent i currently desires that φ be true: $\mathtt{Goal}_i\,\varphi$.

We interpret the second factor, as: agent i believes that there is at least one world, in all possible future worlds, in which φ will be true. In other terms, agent i believes that φ will be true in some possible future worlds: $\mathtt{Poss}_i\mathrm{F}\varphi$. Actually, if i believed that φ can never be the case in all of possible future worlds, then i would have no ground for hope.

The same analysis applies to *Fear*, except that φ is now undesirable for agent i, which we assume that i desires for $\neg\varphi$: $\mathtt{Goal}_i\,\neg\varphi$. We accordingly formalize the cognitive structure of *Hope* and *Fear*:

Definition 2. *(Hope/Fear)*

$$\mathtt{Hope}_i\,\varphi \stackrel{def}{=} \mathtt{Goal}_i\,\varphi \wedge \mathtt{Poss}_i\mathrm{F}\varphi$$
$$\mathtt{Fear}_i\,\varphi \stackrel{def}{=} \mathtt{Goal}_i\,\neg\varphi \wedge \mathtt{Poss}_i\mathrm{F}\varphi$$

For instance, a debutante is hopeful about being asked to dance, if and only if she thinks it is possible ($\mathtt{Poss}_{girl}\mathrm{F}\mathit{be\_asked\_to\_dance}$) and this is what she wants ($\mathtt{Goal}_{girl}\,\mathit{be\_asked\_to\_dance}$). In contrast, an employee fears to be fired when he does not wish to be fired ($\texttt{Goal}_{employee}\,\neg be\_fired$) but believes it is a possibility ($\texttt{Poss}_{employee}\texttt{F}be\_fired$).

Someone now could ask us to add the possibility of $\neg\varphi$ into the formalization of prospect of proposition φ. It is not necessary! Because we adopt the analysis of Ortony et al. [186] who argue that with a prospect proposition φ, if we focused on the possibility of φ, we would hope that φ is the case; if we focused on the possibility of $\neg\varphi$, we would fear about φ not be the case.

Moreover, if we added the possibility of $\neg\varphi$, our formulas would become:

$$\texttt{Hope}_i\,\varphi \overset{def}{=} \texttt{Goal}_i\,\varphi \wedge \texttt{Poss}_i\texttt{F}\varphi \wedge \texttt{Poss}_i\texttt{F}\neg\varphi$$
$$\texttt{Fear}_i\,\varphi \overset{def}{=} \texttt{Goal}_i\,\neg\varphi \wedge \texttt{Poss}_i\texttt{F}\varphi \wedge \texttt{Poss}_i\texttt{F}\neg\varphi$$

this implies $\texttt{Hope}_i\,\varphi \leftrightarrow \texttt{Fear}_i\,\neg\varphi$. This means that one formalization could represent both hope and fear at a time, it is not the hope alone. We therefore could not distinguish these two emotions.

5.1.4 Confirmation-based emotions

5.1.4.1 Satisfaction/Disappointment

Ortony et al. [186] argue that we feel satisfaction when we are pleased about the confirmation of the prospect of a desirable event. Thus 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 two first factors mean that now, agent i keeps in mind that at the previous instant, he desired φ and believed that φ could be true in the future ($\texttt{Bel}_i\,\texttt{X}^{-1}(\texttt{Goal}_i\,\varphi \wedge \texttt{Poss}_i\texttt{F}\varphi)$) (cf. the analysis of the second factor of *Hope*). The last factor means that agent i now experiences that φ is true, but did not know it in the previous instant ($\texttt{Aware}_i\varphi$).

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 ($\mathtt{Bel}_i\, \mathtt{X}^{-1}(\mathtt{Goal}_i\, \neg\varphi \wedge \mathtt{Poss}_i \mathtt{F}\neg\varphi)$). We accordingly formalize the cognitive structure of *Satisfaction* and *Disappointment*:

Definition 3. *(Satisfaction/Disappointment)*

$$\mathtt{Satisfaction}_i\, \varphi \stackrel{def}{=} \mathtt{Bel}_i\, \mathtt{X}^{-1}(\mathtt{Goal}_i\, \varphi \wedge \mathtt{Poss}_i \mathtt{F}\varphi) \wedge \mathtt{Aware}_i \varphi$$

$$\mathtt{Disappointment}_i\, \varphi \stackrel{def}{=} \mathtt{Bel}_i\, \mathtt{X}^{-1}(\mathtt{Goal}_i\, \neg\varphi \wedge \mathtt{Poss}_i \mathtt{F}\neg\varphi) \wedge \mathtt{Aware}_i \varphi$$

For instance, when the debutante realizes that she is indeed asked to dance ($\mathtt{Aware}_{girl} be\_asked\_to\_dance$) she is satisfied. Were she not to be asked to dance ($\mathtt{Aware}_{girl} \neg be\_asked\_to\_dance$), she would feel disappointment.

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

$$\mathtt{Satisfaction}_i\, \varphi \leftrightarrow \mathtt{Bel}_i\, \mathtt{X}^{-1} \mathtt{Hope}_i\, \varphi \wedge \mathtt{Aware}_i \varphi \tag{5.1}$$

$$\mathtt{Disappointment}_i\, \varphi \leftrightarrow \mathtt{Bel}_i\, \mathtt{X}^{-1} \mathtt{Hope}_i\, \neg\varphi \wedge \mathtt{Aware}_i \varphi \tag{5.2}$$

The relation between *Satisfaction* and *Joy*, and that between *Disappointment* and *Distress* can be formalized as Prop. 4a and Prop. 4b, respectively: if we feel satisfaction about something, then we will also feel joy about it; and contrary, if we feel disappointment about something, then we will also feel distress about it.

Proposition 4.

$$\mathtt{Satisfaction}_i \varphi \rightarrow \mathtt{Joy}_i \varphi \tag{Prop. 4a}$$

$$\mathtt{Disappointment}_i \varphi \rightarrow \mathtt{Distress}_i \varphi \tag{Prop. 4b}$$

Proof of Prop. 4a:

$$\mathtt{Satisfaction}_i \varphi \rightarrow \mathtt{Bel}_i\, \mathtt{X}^{-1}(\mathtt{Goal}_i\, \varphi \wedge \mathtt{Poss}_i \mathtt{F}\varphi) \wedge \mathtt{Aware}_i \varphi$$

(From Def. 3)

$$\rightarrow \mathtt{Bel}_i\, \mathtt{X}^{-1} \mathtt{Goal}_i\, \varphi \wedge \mathtt{Aware}_i \varphi$$

$$\rightarrow \mathtt{Joy}_i \varphi \qquad \text{(from Def. 1)}$$

Proof of Prop. 4b:

$$\begin{aligned}
\mathtt{Disappointment}_i\varphi \rightarrow & \mathtt{Bel}_i\,\mathtt{X}^{-1}(\mathtt{Goal}_i\,\neg\varphi \wedge \mathtt{Poss}_i\mathtt{F}\varphi) \wedge \mathtt{Aware}_i\varphi \\
& \hfill \text{(From Def. 3)} \\
\rightarrow & \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Goal}_i\,\neg\varphi \wedge \mathtt{Aware}_i\varphi \\
\rightarrow & \mathtt{Distress}_i\varphi \qquad \text{(from Def. 1)}
\end{aligned}$$

□

This proposition also covers the idea of Ortony et al. [186] who argue that when a desirable proposition φ holds, if we did not focus on the prospect of φ beforehand, we would not feel satisfactory, we would feel only joy. In the contrary, when a desirable proposition φ is not the case, if we did not focus on the prospect of φ beforehand, we would not feel disappointment, we would feel only distress.

5.1.4.2 Fear-confirmed/Relief

Ortony et al. [186] argue that we feel fear-confirmed when we are displeased about the confirmation of the prospect of an undesirable event. Thus the cognitive structure of *Fear-confirmed* consists of three factors:

i. A proposition φ was undesirable for agent i,

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 true.

We use the same analysis as for *Satisfaction*, except in the two first factors, agent recalls that in the previous instant, he undesired the proposition φ; which we assume that agent i desired $\neg\varphi$ ($\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Goal}_i\,\neg\varphi$).

The difference in the case of *Relief* is agent recalls that, in the previous instant, he desired for the proposition φ ($\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Goal}_i\,\varphi$), and believed that $\neg\varphi$ might be true in the near future ($\mathtt{Bel}_i\,\mathtt{X}^{-1}(\mathtt{Goal}_i\,\varphi \wedge \mathtt{Poss}_i\mathtt{F}\neg\varphi)$). We accordingly formalize the cognitive structure of *Fear-confirmed* and *Relief*:

Definition 5. *(Fear-confirmed/Relief)*

$$\mathtt{FearConfirmed}_i\,\varphi \stackrel{def}{=} \mathtt{Bel}_i\,\mathtt{X}^{-1}(\mathtt{Goal}_i\,\neg\varphi \wedge \mathtt{Poss}_i\mathtt{F}\varphi) \wedge \mathtt{Aware}_i\varphi$$
$$\mathtt{Relief}_i\,\varphi \stackrel{def}{=} \mathtt{Bel}_i\,\mathtt{X}^{-1}(\mathtt{Goal}_i\,\varphi \wedge \mathtt{Poss}_i\mathtt{F}\neg\varphi) \wedge \mathtt{Aware}_i\varphi$$

For instance, the employee's fear of being fired is confirmed when he is aware that he is indeed about to be fired ($\mathtt{Aware}_{employee} be\_fired$) which he had been afraid of ($\mathtt{Bel}_{employee}\,\mathtt{X}^{-1}(\mathtt{Goal}_{employee}\,\neg be\_fired \wedge \mathtt{Poss}_{employee}\mathtt{F} be\_fired)$). In the contrary, were he aware that he is not going to be fired ($\mathtt{Aware}_{employee}\neg be\_fired$), he would feel relief.

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

$$\mathtt{FearConfirmed}_i\,\varphi \leftrightarrow \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Fear}_i\,\varphi \wedge \mathtt{Aware}_i\varphi \tag{5.3}$$
$$\mathtt{Relief}_i\,\varphi \leftrightarrow \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Fear}_i\,\neg\varphi \wedge \mathtt{Aware}_i\varphi \tag{5.4}$$

The relation between *Fear-confirmed* and *Distress*, and that between *Relief* and *Joy* are stated in Prop. 6a and Prop. 6b, respectively: if our fears about something are confirmed, then we feel distressed; and in the contrary, if we feel relief from a fear, we also feel joyful about it.

Proposition 6.

$$\mathtt{FearConfirmed}_i\,\varphi \rightarrow \mathtt{Distress}_i\,\varphi \tag{Prop. 6a}$$
$$\mathtt{Relief}_i\,\varphi \rightarrow \mathtt{Joy}_i\varphi \tag{Prop. 6b}$$

Proof of Prop. 6a:

$$\mathtt{FearConfirmed}_i\varphi \rightarrow \mathtt{Bel}_i\,\mathtt{X}^{-1}(\mathtt{Goal}_i\,\neg\varphi \wedge \mathtt{Poss}_i\mathtt{F}\varphi) \wedge \mathtt{Aware}_i\varphi \qquad \text{(From Def. 5)}$$
$$\rightarrow \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Goal}_i\,\neg\varphi \wedge \mathtt{Aware}_i\varphi$$
$$\rightarrow \mathtt{Distress}_i\varphi \qquad \text{(from Def. 1)}$$

Proof of Prop. 6b:

$$
\begin{aligned}
\mathtt{Relief}_i\,\varphi \rightarrow & \mathtt{Bel}_i\,\mathtt{X}^{-1}(\mathtt{Goal}_i\,\varphi \wedge \mathtt{Poss}_i\mathtt{F}\neg\varphi) \wedge \mathtt{Aware}_i\varphi && \text{(From Def. 5)}\\
\rightarrow & \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Goal}_i\,\varphi \wedge \mathtt{Aware}_i\varphi &&\\
\rightarrow & \mathtt{Joy}_i\varphi && \text{(from Def. 1)}
\end{aligned}
$$

□

This proposition also covers the idea of Ortony et al. [186] who argue that when an undesirable proposition φ holds, if we did not focus on the prospect of φ beforehand, we would not feel fear-confirmed, we would feel only distress. In the contrary, when a desirable proposition φ holds, if we did not focus on the prospect of φ beforehand, we would not feel relief, we would feel only joy.

5.1.5 Compound emotions

5.1.5.1 Gratitude

Generally, we are grateful when someone helps us to do something. For instance, Lazarus [135, 136], McCullough et al. [161], Bartlett and DeSteno [14] argue that the receipt of another's help often elicits gratitude. Similarly, Roseman [209, 210] and Weiner [255] also argue that gratitude arises when individuals recognize that they have received a positive outcome, and the positive outcome is contributed from the efforts of another. Ortony et al. [186] argue that gratitude is an approving of someone else's praiseworthy action and being pleased about the related desirable event. We interpret, in a large meaning, that a praiseworthy action is the one which brings a good will outcome to its objective. We accordingly summarize that the cognitive structure of *Gratitude* consists of two factors:

i. A proposition φ was desirable for agent i just before the execution of an action α by agent j, and

ii. agent i believes that agent j has done action α whose outcome is φ, and i experiences that φ is true.

Someone could think about some counter example for our analysis. For instance, at a dinner, a man asks his friend to give to him a pepper, and then his friend does it for him. Does the man feel grateful about his friend's action? The answer is yes, because the man will say "thanks!" to his friend. The difference between this situation with the situation in which a stranger saved the life of a drowning child is only the intensity of being grateful: the more the action is difficult and the goal is strong, the stronger the gratitude is. However, we do not distinguish the intensity of emotion here, we thus consider both situations as having the presence of gratitude.

We so interpret the first factor as: $\mathtt{Bel}_i\, \mathtt{Done}_{j:\alpha}\mathtt{Goal}_i\, \varphi$, and the second factor: agent i believes that φ is outcome of action α ($\mathtt{Bel}_i\, \mathtt{After}_{j:\alpha}\varphi$) and now, agent i experiences that in fact, the expected outcome becomes true ($\mathtt{Aware}_i\varphi$). Thus, we accordingly formalize the cognitive structure of *Gratitude* as:

Definition 7 (Gratitude)**.**

$$\begin{aligned}\mathtt{Gratitude}_{i,j}(\alpha,\varphi) \stackrel{def}{=} & \mathtt{Bel}_i\, \mathtt{Done}_{j:\alpha}\mathtt{Goal}_i\, \varphi \wedge \\ & \mathtt{Bel}_i\, \mathtt{After}_{j:\alpha}\varphi \wedge \mathtt{Aware}_i\varphi\end{aligned}$$

For instance, a mother is grateful that a stranger saved the life of her drowning child because she desires that her child's life be saved ($\mathtt{Goal}_{woman}\, son\_is\_saved$), and because she experiences that her child's life is indeed saved ($\mathtt{Aware}_{woman} son\_is\_saved$).

5.1.5.2 Anger

Generally, we are angry when we feel being slighted or hurt by someone else. For instance, Frijda [81] argues that anger is evoked by being slighted or hurt, or perceiving to be slighted or hurt. Elster [69] also argues that anger is triggered by an undeserved slight or belittling. Sanders [218] pointed out that we become angry at other if we think they did something wrong, we did not want them to do it, and we think they knew it was wrong. Ortony et al. [186] argue that anger is a disapproving of someone else's blameworthy action and being displeased about the related undesirable event. We interpret, in a large meaning, that a blameworthy action is the one which brings an undesirable (bad will) outcome to its objective. We accordingly summarize that the cognitive structure of *Anger* consists of two factors:

i. A proposition φ was undesirable for agent i just before the execution of an action α by agent j, and

ii. agent i believes that agent j has done action α whose outcome is φ, and i experiences that the undesirable proposition φ is true.

Someone now could think about some counter example of our analysis. For instance, a player plays in a sportive competition of tennis whose rule enables only one player to win, and he lost. In fact, this is an complex situation that can bring different emotions in a time, in particular disappointment and anger. If the man focused on his opponent or the referee (ex. his opponent cheated or the referee has mistaken) as the reason why he lost the game, he would feel angry about the opponent or the referee (this case corresponds to our analysis). But if the man focused only on the lack of his competence regarding his opponent's talent in playing the tennis, he would feel only disappointment: he was hopeful to win, so he feels disappointment instead of anger when he lost.

We so interpret the first factor as: $\texttt{Bel}_i\, \texttt{Done}_{j:\alpha}\texttt{Goal}_i\, \neg\varphi$; and the second factor: agent i believes that φ is the outcome of action α ($\texttt{Bel}_i\, \texttt{After}_{j:\alpha}\varphi$) and now agent i experiences that in fact the unexpected outcome becomes true ($\texttt{Aware}_i\varphi$). We accordingly formalize the cognitive structure of *Anger* as:

Definition 8 (Anger)**.**

$$\begin{aligned}\texttt{Anger}_{i,j}(\alpha,\varphi) \overset{def}{=} & \texttt{Bel}_i\, \texttt{Done}_{j:\alpha}\texttt{Goal}_i\, \neg\varphi \wedge \\ & \texttt{Bel}_i\, \texttt{After}_{j:\alpha}\varphi \wedge \texttt{Aware}_i\varphi\end{aligned}$$

For instance, you would be angry about being robbed by a burglar while being out of the house. The anger ($\mathtt{Anger}_{you,burglar}(robbery, lost\_possessions)$) because you wished to keep your possessions ($\mathtt{Goal}_{you}\,\neg lost\_possessions$); now a burglar robbed you, and you experience the loss of your possessions ($\mathtt{Aware}_{you} lost_p ossessions$).

Anger in inaction

In the case of anger in inaction, $\mathtt{Anger}_{i,j}(\sim \alpha, \varphi)$, reads "agent i is angry about agent j to not do action α to avoid φ". In this case, φ could be avoided by doing action α, we then interpret it as if j performs α, we will have $\neg\varphi$ instead of φ. So we accordingly have the formula of the cognitive structure of anger in inaction:

$$\begin{aligned}\mathtt{Anger}_{i,j}(\sim \alpha, \varphi) \overset{def}{=} & \mathtt{Bel}_i\, \mathtt{Done}_{\sim j:\alpha} \mathtt{Goal}_i\, \neg\varphi \wedge \\ & \mathtt{Bel}_i\, \mathtt{After}_{j:\alpha} \neg\varphi \wedge \mathtt{Aware}_i \varphi\end{aligned} \tag{5.5}$$

5.1.6 Discussion

In this section, we discuss how does our model respect the nature features of emotion such as subjectivity, instant nature and intensity. We also discuss about the compatibility of our model regarding other recent proposed models.

5.1.6.1 Subjectivity of emotion

Emotion is subjective. It depends on what and how its owner receipt about a related event. Let consider *Joy*, for instance, which occurs when someone experienced that a desirable proposition be true. The first factor relates to a desirable proposition φ: if agent i did not know or did forget his desire, he would not feel joyful when the related event occurs. The same for the second factor which is related to the experience of event φ: if the event did occur but the agent do not know it, he would not be joyful too. In the contrary, if the event did not occur but the agent believed that it happened (a wrong belief), he would be joyful, even then his joy will be disappeared when the truth is reviewed.

Our model respects the subjectivity of emotion by explicitly putting the $\mathtt{Bel}_i$ operator before almost cognitive factors of each emotion. For instance of *Joy*: $\mathtt{Joy}_i\,\varphi \overset{def}{=} \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Goal}_i\,\varphi \wedge \mathtt{Aware}_i\varphi$, the first component means that agent i recalls his desire from previous instant, the second component means that agent now experiences that φ just happened in order to be joyful.

This respect has only two exceptions in the definition of *Hope* ($\mathtt{Hope}_i\,\varphi \overset{def}{=} \mathtt{Goal}_i\,\varphi \wedge \mathtt{Poss}_i\mathtt{F}\varphi$) and *Fear* ($\mathtt{Fear}_i\,\varphi \overset{def}{=} \mathtt{Goal}_i\,\neg\varphi \wedge \mathtt{Poss}_i\mathtt{F}\varphi$) because we do not need to explicitly put $\mathtt{Bel}_i$ before each component. From the semantic of our logic, this is implicitly equivalent to the case of putting it. Regarding the first component of *Hope*, from the definition of the abbreviation $\mathrm{Def}_{\mathtt{Goal}_i}$: $\mathtt{Goal}_i\,\varphi \overset{def}{=} \mathtt{Choice}_i\,\mathtt{F}\mathtt{Bel}_i\,\varphi$, and from the axiom 4_{BC} $\mathtt{Choice}_i\,\varphi \leftrightarrow \mathtt{Bel}_i\,\mathtt{Choice}_i\,\varphi$ (agent is conscious about his chosen preferences), we will have $\mathtt{Goal}_i\,\varphi \leftrightarrow \mathtt{Bel}_i\,\mathtt{Goal}_i\,\varphi$. The same analysis for the second component of *Hope*, from the definition of the abbreviation $\mathrm{Def}_{\mathtt{Poss}_i}$: $\mathtt{Poss}_i\varphi \overset{def}{=} \neg\mathtt{Bel}_i\,\neg\varphi$, and from the axiom $5_{\mathtt{Bel}_i}$: $\neg\mathtt{Bel}_i\,\varphi \leftrightarrow \mathtt{Bel}_i\,\neg\mathtt{Bel}_i\,\varphi$ (agent is conscious about his disbeliefs), we will have $\mathtt{Poss}_i\varphi \leftrightarrow \mathtt{Bel}_i\,\mathtt{Poss}_i\varphi$. That is why we do not need to explicitly put the $\mathtt{Bel}_i$ before components which begin with $\mathtt{Goal}_i$ or $\mathtt{Poss}_i$. This rule will be also applied for the formalization of concept of trust and distrust in the section 5.2.

5.1.6.2 Instant nature of emotion

As mentioned in the analysis of *Joy/Distress*, emotion has an instant nature: it does not exist forever. We respect this feature of emotion by using the operator $\mathtt{Aware}_i\varphi$ in almost emotions, except *Hope* and *Fear* which have a long duration until the prospect event occurs.

Let us show how this operator avoids the case that emotion exists forever. For instance, if we feel joy about φ now, it means that $\mathtt{Aware}_i\varphi$ is currently true, then neither $\mathtt{Aware}_i\varphi$ was true in previous instant because (from the

definition of the abbreviation $\mathrm{Def}_{\mathtt{Aware}_i}$)
$\mathtt{X}^{-1}\mathtt{Aware}_i\varphi \wedge \mathtt{Aware}_i\varphi \rightarrow \mathtt{X}^{-1}(\mathtt{Bel}_i\, \varphi \wedge \neg\mathtt{Bel}_i\, \varphi) \rightarrow \bot$
nor will $\mathtt{Aware}_i\varphi$ be true in the next instant because
$\mathtt{Aware}_i\varphi \wedge \mathtt{XAware}_i\varphi \rightarrow \mathtt{Bel}_i\, \varphi \wedge \neg\mathtt{Bel}_i\, \varphi \rightarrow \bot$
So, neither we did feel joyful in previous instant, nor will we feel joyful in the next instant. The instant nature of emotion (*Joy*) is fully respected.

The respect of emotion's instant nature could be seen as one of the main advantages of our model regarding recent models proposed such as the model of Adam et al. [3] in which, for example, *Joy* is defined as $Des_i\varphi \wedge \mathtt{Bel}_i\, \varphi$. If φ not changes after being true, the second component will be true forever. So, if we keep our desire, we will feel joyful forever! It is not intuitive with respecting the nature of emotion.

5.1.6.3 Intensity of emotion

Although we have added time factor into almost emotional formulas, which enables to eliminate rightly emotion when the relevant event has passed, but it has not yet helped us to represent the nature of continuous intensity of emotions.

In our model, emotion is triggered when the related proposition appears and all of its (cognitive) factors are fulfilled, presents in current instant and then, disappears in the next instant. In other terms, emotion has only two states: either *enable* or *disable*. The reason is our operators have also only two boolean values, and there is no intermediate values between *true* and *false*. This seem to be a common inconvenient of all approaches which are based on logic regarding the model of Adam et al. [3], Ochs et al. [181, 182] and the one of Meyer [164]. These models, like ours, mainly focus on qualitative formalization of emotion instead of quantitative aspect. However, these approaches are more helpful to disambiguate the emotion and to the ability of reasoning about emotion.

5.1.6.4 Compatibility with other approaches

Adam et al. [3] based on the same foundation model of emotion as ours, the theory of Ortony et al. [186]. At the logical level, their formal framework is based on the modal logic of belief, choice, time, and action of Herzig and Longin [105] which is a refinement of Cohen and Levesque's works [50] and the BDI modal logic of Georgeff et al. [86]. The difference is on the semantic of time: in their logic, a story is a set of possible worlds plus some properties of the corresponding accessibility relation that entails the linearity of time. While in our logic, a story is a primitive of the semantics. Moreover, we add the next operator $\mathtt{X}$ and converse $\mathtt{X}^{-1}$ to the logical language, and the formalization of emotions are thus different. For instance, *Joy* is defined here as ($\mathtt{Joy}_i\,\varphi \overset{def}{=} \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Goal}_i\,\varphi \land \mathtt{Aware}_i\varphi$), but was defined as ($Des_i\varphi \land \mathtt{Bel}_i\,\varphi$) in [3].

Meyer [164] presents a logic which enables to reason about the dynamics of emotional states of agent. This logic is based on a BDI model [86], and the framework of dynamic logic and (an extension of) the KARO framework [249]. Therefore, they focus on the action in their emotion's definition (original of dynamic logic). This is the main difference from us, but the principle is compatible. Regarding, for instance, their definition of *Happy* (refers to the same thing of our *Joy*):
$I(\pi,\varphi) \land Com(\pi) \rightarrow [\pi](\varphi \rightarrow happy(\epsilon,\varphi,\varphi))$
expressing that an agent that was committed to a plan π which he intended to do with φ as goal, and that believes that his goal φ is realised after having executed/performed his plan π, is indeed happy with this. If we removed the fact of action π, it would remain
$I(\varphi) \rightarrow (\varphi \rightarrow happy(\epsilon,\varphi,\varphi))$
could be read "if we want to have φ then: if we have it, then we feel happy". It is similar to our definition of *Joy* without the time factor!

Ochs et al. [181, 182] also use the theory of Ortony et al. [186] as a foundation model. They also based a logic of "Belief" and "Desire" to construct their emotional model. The main difference from our approach is that they based on the capability of emotion's owner. Moreover, the logic is more used as a language of representation than a tool of reasoning.

Steunebrink et al. [238, 240] also use the theory of Ortony et al. [186] as a foundation model but they formalize all twenty-two emotions of Ortony et al. as primitive operators in order to reasoning how about emotion effects on action. Their model is thus far from ours: each emotion has a different etiquette and there is no common component among emotions. It is also a philosophical point of view that we do not argue. In our work, emotions are based on some mental attitudes and thus they are not primitive of our language.

5.1.6.5 Compatibility with other foundation models

As we mentioned in the choice of foundation model section, our model is mainly based on the theory of Ortony et al. [186], Lazarus [135] and Frijda [81]. However, our model also refers and is compatible to other theories such as theory BDTE of Reisenzein [205].

Reisenzein [205] proposed the Belief-Desire Theory of Emotion (BDTE) with both qualitative and quantitative aspects of emotion. Regarding the qualitative aspect, our model is very compatible to theirs: they use the same notion of Bel_i to represent what agent believes (refers to our $\mathtt{Bel}_i$) and Des_i to represent what agent choose (refers to our $\mathtt{Goal}_i$). Their uncertain of a proposition p could be covered by our $\mathtt{Poss}_i$ operator. For the time aspect, our $\mathtt{X}^{-1}$ could refer to their time $t-1$ as well as our $\mathtt{X}$ could refer to their time $t+1$. However, they define only seven basic emotions (happy/unhappy, hopes/fears, surprised, disappointment/relieve) but there is no definition for action-oriented emotions such as anger and gratitude. Thus, we can not compare with them about these last two emotions.

Note that we do not consider *surprise* as an emotion here. For us, the fact that an emotion has necessarily a valence (i.e. an emotion must be either positive or negative) is a fundamental property of emotion. But surprise may be sometimes positive, sometimes negative. Thus, following this property of valence, it is not an emotion. Ortony et al. [186] agree with that for instance.

5.2 Formalization of Trust & Distrust

5.2.1 Choice of foundation models

In order to formalize trust for rational intelligent agents, we tend to base on the cognitive definitions of trust/distrust which argues that trust/distrust must be conscious. The reasons are the same as those mentioned when we choose the foundation models for formalization of emotion: the cognitive models not only enable us to recognize and reasoning about trust/distrust but also easier to formalize because they are supportable to AI applications.

As a potential model, the theory of Jones [121] considers trust as a belief: it is suggested that two beliefs - called the “rule-belief” and the “conformity-belief” - form the core of the trusting attitude. Jones analyzed five scenarios in which, it is true to say that i trusts j:

- belief in regularity of trustee;
- belief in trustee’s obligation;
- belief in role of trustee;
- belief in trustee’s informing; and
- belief in trustee’s intention.

This seem to be clear and usable. But in our opinion, trust could be betrayed and once being betrayed, trustor will be angry. This analysis is not supported by the trust’s definition of Jones. For instance, a scenario of the belief in regularity of trustee is: A man always observes that his neighbor outs to work on every morning, so he trusts (in the definition of Jones) his neighbor to go out on the morning. But, imagine that one day, the man discovers that his neighbor does not go out on the morning, so could he be angry about his neighbor's action? It is clearly that no anger! So in our opinion, there was no trust of such a man, that was just a kind of reliance (for more details, see Longin et al. [145]).

In our work, we take into account both trust and distrust's cognitive definition of Castelfranchi and colleagues [39, 41, 75] who argue that trust is based on a combination of some kinds of beliefs: competence belief, dependence belief, disposition belief (which is decomposed into willingness belief and persistence belief), self-confidence belief and motivation belief of trustor on trustee. This model seem to be helpful to us to formalize the related concepts because it is popularly used for others works such as those of Herzig et al. [103, 104].

5.2.2 Formalization of Trust

We formalize the concept of trust based on Castelfranchi and Falcone's definition [41, 75] of trust in action which says that agent i trusts agent j to ensure φ by performing action α if and only if:

i. agent i desires to achieve φ, which is formalized as $\texttt{Goal}_i\, \varphi$, and agent i expects that:

ii. φ can be achieved by doing action α, which is interpreted that φ is obtained after every execution of action α: $\texttt{Bel}_i\, \texttt{After}_{j:\alpha}\varphi$;

iii. agent j is able to perform action α which is interpreted as the meaning of the abbreviation $\texttt{Able}_j$: $\texttt{Bel}_i\, \texttt{Able}_j\alpha$; and

iv. agent j has the intention to do such an action which is interpreted as the meaning of the abbreviation $\texttt{Intend}_j$: $\texttt{Bel}_i\, \texttt{Intend}_j\, \alpha$.

However, these three factors are only necessary conditions, but not sufficient ones[2]. For instance, imagine that a robber wants to steal something

[2]The more trust conditions are strong, the stronger is the negative emotion that we feel when trust is betrayed. Thus, we prefer here a stronger trust formalization than that of Herzig et al. [103, 104].

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, all 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 public ground between the nurse (trustee) and the robber (trustor).

So here we need to add another condition for trust: trustor also believes that there is a public ground between trustor and trustee about the fact that the trustee will perform such an action ($\texttt{Bel}_{trustor}\,\texttt{Grd}_I trustee : \alpha\text{-}happens$), where $I = \{trustor, trustee\}$. We put the $\texttt{Bel}_{trustor}$ before the $\texttt{Grd}_I$ to indicate that trust is always subjective belief of trustor. This always respects the subjectivity of trust (cf. Ries et al. [206]). We accordingly formalize the cognitive structure of *Trust* as:

Definition 9. *(Trust)*

$$\begin{aligned}\texttt{Trust}_{i,j}(\alpha,\varphi) \overset{def}{=} &\texttt{Goal}_i\,\varphi \wedge \texttt{Bel}_i\,\texttt{After}_{j:\alpha}\varphi \wedge \texttt{Bel}_i\,\texttt{Able}_j\alpha \wedge \\ &\texttt{Bel}_i\,\texttt{Intend}_j\,\alpha \wedge \texttt{Bel}_i\,\texttt{Grd}_{\{i,j\}} j{:}\alpha\text{-}happens\end{aligned}$$

For instance, a boss trusts his secretary to prepare a report in order to present it at a company meeting because the boss desires the report ($\texttt{Goal}_{boss}\,report$), and in his opinion, the report can be possibly ready after the secretary prepares it ($\texttt{Bel}_{boss}\,\texttt{After}_{secretary:prepare} report$), the secretary has the ability and intention to prepare the report ($\texttt{Bel}_{boss}\,\texttt{Able}_{secretary} prepare \wedge \texttt{Bel}_{boss}\,\texttt{Intend}_{secretary}\,prepare$). Moreover, the boss believes that there is a public ground that the secretary will prepare the report in time ($\texttt{Bel}_{boss}\,\texttt{Grd}_{boss,secretary} secretary : prepare\text{-}happens$).

Trust in inaction

Demolombe and Lorini [58] distinguished trust in the trustee's action from trust in the trustee's inaction: in the former case, truster believes that trustee is in condition to *further* his goal, and trustee will *do* that; in the latter case, truster believes that trustee is in condition to *endanger* his goal, but trustee will *refrain* from doing that. This opposition is symmetrical to the opposition between *doing* and *refraining* which is studied in the philosophy of action (cf. Belnap et al. [18] and van Benthem [247]). According to this analysis, Herzig and Lorini [104] formalized that i trusts j not to do α to bring about φ if only if i wants φ to be true ($\mathtt{Goal}_i\,\varphi$) and i believes that j, by doing α, will bring about $\neg\varphi$ ($\mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\neg\varphi$), j is able to do α ($\mathtt{Bel}_i\,\mathtt{Able}_j\alpha$), but j has no intention to do α ($\mathtt{Bel}_i\,\neg\mathtt{Intend}_j\,\alpha$). We extent this formalization with respecting our analysis that in oder to trust, truster has a belief about a public ground that trustee will *not do* action α ($\mathtt{Bel}_i\,\mathtt{Grd}_{\{i,j\}}\neg j{:}\alpha\text{-}happens$).

We accordingly have the formalization of the cognitive structure of trust in inaction:

$$\mathtt{Trust}_{i,j}(\sim\alpha,\varphi) \overset{def}{=} \mathtt{Goal}_i\,\varphi \wedge \mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\neg\varphi \wedge \mathtt{Bel}_i\,\mathtt{Able}_j\alpha \wedge \mathtt{Bel}_i\,\neg\mathtt{Intend}_j\,\alpha \wedge \mathtt{Bel}_i\,\mathtt{Grd}_{\{i,j\}}\neg j{:}\alpha\text{-}happens \quad (5.6)$$

5.2.3 Formalization of Distrust

We also adopt the definition of distrust given by Castelfranchi et al. [39] which says that agent i distrusts agent j to ensure φ by performing action α if and only if:

i. agent i desires to achieve φ ($\mathtt{Goal}_i\,\varphi$), and agent i believes that at least one of these conditions is fulfilled:

ii. agent j is not able to bring about φ by performing action α: $\mathtt{Bel}_i\,\neg\mathtt{After}_{j:\alpha}\varphi$, or

iii. agent j is able to bring about φ by performing action α but he has not intention to do α: $\mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\varphi \wedge \mathtt{Bel}_i\,\neg\mathtt{Intend}_j\,\alpha$.

We accordingly formalize the cognitive structure of *Distrust* as:

Definition 10. *(Distrust)*

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

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

From this definition, we can decompose the concept of distrust based only upon the lack of trustee's ability and distrust based only upon the lack of trustee's intention:

Definition 11. *(Distrust based on ability)*

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

Definition 12. *(Distrust based on intention)*

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

5.2.4 Discussion

This section compares our formalization with other approaches in formal definition of (individual) trust which are presented in section 3.2.3. In the Table 5.1, the first column represents models' owners, six remain columns correspond to six considered features as follows:

- Cognitive: to see whether mental states like beliefs are used in the definition.
- Goal: to see whether trustor has a goal to trust other on bringing about it.
- Public ground: to see whether trustor believes there is a public ground between him and trustee.
- Action: to see whether trustor focuses on the action which brings about what he trusts.

Model	Trustor			Trustee		
	Cognitive	Goal	Public ground	Action	Competence	Intention
Jones & Firozabadi	✓					✓(goal)
Grandison & Reichgelt(E2T2)	✓					✓(goal)
Herzig et al.	✓	✓		✓	✓	✓
Our model	✓	✓	✓	✓	✓	✓

Table 5.1: Comparison between formal trust's definition approaches

- Competence: to see whether trustor considers the trustee's competence.

- Intention: to see whether trustor considers the trustee's intention.

From the Table 5.1, the first two models, including the model of Jones and Firozabadi [123], and the model E2T2 of Grandison and Reichgelt [92] are quite similar: trust is simply a belief of trustor on what the trustee wants to do. In these models, trustee's intention is assimilated to trustee's goal which is completely different from the trustor's goal mentioned in the third column of the table. They consider neither trustor's goal which trustee is going to do, nor the action which causes the result that trustor want to be the case. Consequently, they do not considers the trustee's competence as a component of their definition.

The most similar to our model is the model of Herzig et al. [103] because they use the same foundation model of trust, it is the definition of trust of Falcone and Castelfranchi [75]; except that in their model, they did not consider the belief of trustor about a public ground between him and trustee as one of condition to trust. As a consequence, their model covers also the case of reliance (not rightly trust) as we discussed in the analysis to formalize trust.

In our model, we add trustor's belief about a public ground between him and trustee as one of condition to trust. Thus, we could consider our model is a tighter case of Herzig et al.'s formalization.

5.3 Conclusion

Using the logic introduced in chapter 4, this chapter presented the formalization of the cognitive structure of emotions, including the well-being emotions as joy and distress; prospect-based emotions as hope and fear; confirmation-based emotions as satisfaction, disappointment, fear-confirmed and relief; and compound emotions as gratitude and anger. This chapter also presented the formalization of the cognitive structure of trust and distrust in the same logic.

From these formal representations of the cognitive structure of emotions, trust and distrust, we will formalize the relations between them, particularly the effect of trust/distrust on emotions, which will be presented in chapter 6.

Although our formalization the cognitive structure of trust is mainly based on the definition of Falcone and Castelfranchi [75] which argues that trust is a combination of trustor's beliefs about trustee's capacity, power and intention, we considered that trust is also composed of trustor's belief about a public ground between him and trustee. We will collect empirical data to validate this argumentation in chapter 7 (section 7.2).

Chapter 6

Effect of Trust/Distrust on Emotions

From the formal representation of cognitive structure of emotions, trust and distrust which are presented in chapter 5, this chapter show that there are some relations between them, particularly the effect of trust/distrust on some emotions. These relations are also represented in the same logic introduced in chapter 4.

Section 6.1 presents the effect of trust on emotions as hope, satisfaction, disappointment and anger; Section 6.2 presents the effect of distrust on emotions as fear, fear-confirmed, relief and gratitude; Section 6.3 discusses about these formalizations.

6.1 Effect of Trust on Emotions

This section presents the formalization of the effect of trust on emotions: *trust* implies *hope* (Proposition 13); a success following trust brings about *satisfaction* (Proposition 14) and/or *gratitude* (Proposition 15); and a failure following trust brings about *disappointment* (Proposition 16) and/or *anger* (Proposition 17).

6.1.1 Trust and Hope

Trust and *Hope* have an important relation because they both feature a positive expectation (cf. Castelfranchi and colleagues [75, 165]). 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 (cf. Bryce [32]). We accordingly formalize the relation between trust and hope as:

Proposition 13. *(Trust implies Hope)*

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

This means that when we trust someone about an action that will bring some results, we are hopeful that the action's results will be obtained. For instance, when a boss trusts his secretary to prepare an urgent report ($\mathtt{Trust}_{boss,secretary}(prepare, report)$) for a meeting in tomorrow, he will hope to have the report on time ($\mathtt{Hope}_{boss}report$).

This proposition will be proved by applying Lemma 13.1: 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 13.1.

$$\mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\varphi \wedge \mathtt{Bel}_i\,\mathtt{Able}_j\alpha \rightarrow \mathtt{Poss}_i\mathtt{F}\varphi$$

Proof.

$$\begin{aligned}
\mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\varphi \rightarrow & \mathtt{Bel}_i\,(j{:}\alpha\text{-}happens \rightarrow \mathtt{X}\varphi) && (\text{from Def}_{\mathtt{After}_{i:\alpha}})\\
\rightarrow & (\mathtt{Bel}_i\,j{:}\alpha\text{-}happens \rightarrow \mathtt{Bel}_i\,\mathtt{X}\varphi) && (\text{from axiom K}_{\mathtt{Bel}_i})
\end{aligned} \tag{6.1}$$

$$\begin{aligned}
\mathtt{Bel}_i\,\mathtt{Able}_j\alpha \rightarrow & \mathtt{Bel}_i\,\neg\mathtt{After}_{j:\alpha}\bot && (\text{from Def}_{\mathtt{Able}_i})\\
\rightarrow & \mathtt{Bel}_i\,\neg(j{:}\alpha\text{-}happens \rightarrow \mathtt{X}\bot) && (\text{from Def}_{\mathtt{After}_{i:\alpha}})\\
\rightarrow & \mathtt{Bel}_i\,j{:}\alpha\text{-}happens
\end{aligned} \tag{6.2}$$

$$\begin{aligned}
\mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\varphi \wedge \mathtt{Bel}_i\,\mathtt{Able}_j\alpha & \\
\rightarrow & (\mathtt{Bel}_i\,j{:}\alpha\text{-}happens \rightarrow \mathtt{Bel}_i\,\mathtt{X}\varphi)\wedge\\
& \mathtt{Bel}_i\,j{:}\alpha\text{-}happens && (\text{from (6.1) and (6.2)})\\
\rightarrow & \mathtt{Bel}_i\,\mathtt{X}\varphi\\
\rightarrow & \mathtt{Poss}_i\mathtt{X}\varphi && (\text{from axiom D}_{\mathtt{Bel}_i})\\
\rightarrow & \mathtt{Poss}_i\mathtt{F}\varphi
\end{aligned}$$

□

Proof of Proposition 13.

$$\mathtt{Trust}_{i,j}(\alpha,\varphi) \rightarrow \mathtt{Goal}_i\,\varphi \quad (\text{from Def. 9}) \tag{6.3}$$

$$\begin{aligned}
\mathtt{Trust}_{i,j}(\alpha,\varphi) \rightarrow & \mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\varphi \wedge \mathtt{Bel}_i\,\mathtt{Able}_j\alpha && (\text{from Def. 9})\\
\rightarrow & \mathtt{Poss}_i\mathtt{F}\varphi && (\text{from Lem. 13.1})
\end{aligned} \tag{6.4}$$

$$\begin{aligned}
\mathtt{Trust}_{i,j}(\alpha,\varphi) \rightarrow & \mathtt{Goal}_i\,\varphi \wedge \mathtt{Poss}_i\mathtt{F}\varphi && (\text{from (6.3) and (6.4)})\\
\rightarrow & \mathtt{Hope}_i\varphi && (\text{from Def. 2})
\end{aligned}$$

□

This proposition is also consistent with the argumentation of Miceli and Castelfranchi [165] who argue that hope and trust do not completely overlap: Trust appears to be stronger than hope. In other term, if there is trust, there will be hope. But the inverse is not correct: if there is hope, we could not be sure that there will be trust.

6.1.2 Success following Trust

A success following the trust of agent i on agent j in doing action α to bring about φ is the fact that agent i finally experiences in fact φ is true. We assume that in this situation, agent i could feel satisfaction:

Proposition 14. *(Success following Trust implies Satisfaction)*

$$\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Trust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\varphi \rightarrow \mathtt{Satisfaction}_i\varphi$$

Proof.

$$\begin{aligned}
\mathtt{Trust}_{i,j}(\alpha,\varphi) &\rightarrow \mathtt{Hope}_i\varphi && \text{(from Prop. 13)}\\
\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Trust}_{i,j}(\alpha,\varphi) &\rightarrow \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Hope}_i\varphi\\
\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Trust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\varphi &\rightarrow \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Hope}_i\varphi \wedge \mathtt{Aware}_i\varphi\\
&\rightarrow \mathtt{Satisfaction}_i\varphi && \text{(from (5.1))}
\end{aligned}$$

□

For instance, once the boss trusted his secretary to prepare the report ($\mathtt{X}^{-1}\mathtt{Trust}_{boss,secretary}(prepare, report)$), and on the morning of the day after, he has the report on his official table ($\mathtt{Aware}_{boss}report$), then he is satisfied ($\mathtt{Satisfaction}_{boss}report$).

This proposition has a corollary: When we experience that what we trusted has really been true, we will also feel joy about it.

Corollary 14.1.

$$\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Trust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\varphi \rightarrow \mathtt{Joy}_i\varphi$$

Proof.

$$\begin{aligned}
\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Trust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\varphi &\rightarrow \mathtt{Satisfaction}_i\varphi\\
&&& \text{(from Prop. 14)}\\
&\rightarrow \mathtt{Joy}_i\varphi && \text{(from Prop. 4a)}
\end{aligned}$$

□

Once a success follows trust, if trustor focuses on the action's owner, trustor will feel gratitude about trustee. We accordingly formalize this relation as:

Proposition 15. *(Success following Trust implies Gratitude)*

$$\mathtt{Bel}_i\,\mathtt{Done}_{j:\alpha}\mathtt{Trust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\varphi \rightarrow \mathtt{Gratitude}_{i,j}(\alpha,\varphi)$$

Proof. In dynamic logic, the formula $[\alpha]\varphi$ reads "after every execution of α, φ is the case". We interpret it as a time independent fact: $[\alpha]\varphi \rightarrow \mathtt{G}[\alpha]\varphi$ for a particular case of $\mathtt{After}_{j:\alpha}\varphi$, we thus have:

$$\mathtt{After}_{j:\alpha}\varphi \rightarrow \mathtt{X}\mathtt{After}_{j:\alpha}\varphi \tag{6.5}$$

$$\mathtt{Trust}_{i,j}(\alpha,\varphi) \rightarrow \mathtt{Goal}_i\,\varphi \wedge \mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\varphi \qquad \text{(from Def. 9)}$$

$$\begin{aligned}
\mathtt{Bel}_i\,\mathtt{Done}_{j:\alpha}\mathtt{Trust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Bel}_i\,\varphi \rightarrow& \\
&\mathtt{Bel}_i\,\mathtt{Done}_{j:\alpha}(\mathtt{Goal}_i\,\varphi \wedge \mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\varphi) \wedge \mathtt{Aware}_i\varphi \\
&\rightarrow \mathtt{Bel}_i\,\mathtt{Done}_{j:\alpha}\mathtt{Goal}_i\,\varphi \wedge \mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\varphi \wedge \mathtt{Aware}_i\varphi \\
&\qquad \text{(from (6.5))} \\
&\rightarrow \mathtt{Gratitude}_{i,j}(\alpha,\varphi) \qquad \text{(from Def. 7)}
\end{aligned}$$

□

To summarize, when a success follows trust, if trustor focuses only on the action's outcome, s/he will feel satisfaction. If trustor also focuses on the action's owner, s/he will feel gratitude about trustee.

6.1.3 Failure following Trust

A failure following the trust of agent i on agent j in doing action α to bring about φ is the fact that agent i finally experiences that $\neg\varphi$ is true instead of φ. We assume that in this situation, agent i could feel disappointment:

Proposition 16. *(Failure following Trust implies Disappointment)*

$$\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Trust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\neg\varphi \rightarrow \mathtt{Disappointment}_i\neg\varphi$$

Proof.

$$
\begin{aligned}
\mathtt{Trust}_{i,j}(\alpha,\varphi) &\rightarrow \mathtt{Hope}_i\varphi \qquad \text{(from Prop. 13)}\\
\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Trust}_{i,j}(\alpha,\varphi) &\rightarrow \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Hope}_i\varphi\\
\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Trust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\neg\varphi &\rightarrow \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Hope}_i\varphi \wedge \mathtt{Aware}_i\neg\varphi\\
&\rightarrow \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Hope}_i\neg(\neg\varphi) \wedge \mathtt{Aware}_i(\neg\varphi)\\
&\rightarrow \mathtt{Disappointment}_i\neg\varphi\\
&\qquad \text{(from (5.2))}
\end{aligned}
$$

□

For instance, the day after, the boss discoveries there is not report on his official table ($\mathtt{Aware}_{boss}\neg report$), he will be disappointed ($\mathtt{Disappointment}_{boss}\neg report$).

This proposition has a corollary: When we experience that what we trusted has really been failed, we will also feel distress about it.

Corollary 16.1.

$$\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Trust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\neg\varphi \rightarrow \mathtt{Distress}_i\neg\varphi$$

Proof.

$$
\begin{aligned}
\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Trust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\neg\varphi &\rightarrow \mathtt{Disappointment}_i\neg\varphi\\
&\qquad \text{(from Prop. 16)}\\
&\rightarrow \mathtt{Distress}_i\neg\varphi \qquad \text{(from Prop. 4b)}
\end{aligned}
$$

□

As having presented in section 3.1.3, betrayal is a feature of trust. The trust of agent i on agent j in doing (resp. avoiding) action α to bring about φ is betrayed when (i) agent i knows that agent j did not do action α (resp. did it), and (ii) agent i experiences that in fact $\neg\varphi$ is true instead of φ. We assume that in this situation, agent i could be angry. We accordingly formalize this relation as:

Proposition 17. *(Betrayal of Trust implies Anger)*

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

Proof.

$$\mathtt{Trust}_{i,j}(\sim\alpha,\varphi)\rightarrow\mathtt{Goal}_i\,\varphi\wedge\mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\neg\varphi \qquad \text{(from (5.6))}$$

$$\mathtt{Done}_{j:\alpha}\mathtt{Trust}_{i,j}(\sim\alpha,\varphi)\wedge\mathtt{Aware}_i\neg\varphi\rightarrow$$

$$\mathtt{Done}_{j:\alpha}(\mathtt{Goal}_i\,\varphi\wedge\mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\neg\varphi)\wedge\mathtt{Aware}_i\neg\varphi$$

$$\rightarrow\mathtt{Done}_{j:\alpha}\mathtt{Goal}_i\,\varphi\wedge\mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\neg\varphi\wedge\mathtt{Aware}_i\neg\varphi \qquad \text{(from (6.5))}$$

$$\rightarrow\mathtt{Anger}_{i,j}(\alpha,\neg\varphi) \qquad \text{(from Def. 8)}$$

□

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

For instance, Bob and John are two closed friends. In fact, Bob was in prison because he hit a girl. Bob wants to hide his secret. Thus, he did not tell anyone his secret except John because Bob trusted John to not reveal his secret to anyone ($\mathtt{X}^{-1}\mathtt{Trust}_{Bob,John}(\neg reveal, secret\_is\_hiden)$). One day, Bob discovers that John has spoken to someone his secret ($\mathtt{Aware}_{Bob}\neg secret\_is\_hiden$). So he is angry about John's action ($\mathtt{Anger}_{Bob,John}(reveal, \neg secret\_is\_hiden)$).

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

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

This formula corresponds to the case of anger in inaction which comes from trust in action. Thus, based on the idea that anger comes from the betrayal of trust, Proposition 17 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.

To summarize, when a failure follows trust, if trustor focuses only on the action's outcome, s/he will feel disappointment. If trustor also focuses on the action's owner, s/he will feel anger because of being betrayed.

6.2 Effect of DisTrust on Emotions

This section presents the formalization of the effect of distrust on emotions: *distrust* implies *fear* (proposition 18); a failure following distrust brings about *fear-confirmed* (proposition 19); and a success followings distrust brings about *relief* (proposition 20) and/or *gratitude* (proposition 21).

6.2.1 DisTrust and Fear

Distrust features a negative expectation, involving fear of the other (Lewicki and Wiethoff [139], Aghio et al. [6]). Castelfranchi and Falcone also argue that fear is a possible part of distrust [41, p. 122]. We state the relation between *Distrust* based on ability and *Fear* as Proposition 18.

Proposition 18. *(DisTrust implies Fear)*

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

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 instance, the trusted secretary has some other business, thus the boss has to ask his assistant to prepare the report he needs in spite that the boss distrusts his assistant to finish the report by the next morning ($\texttt{C-DisTrust}_{boss,assistant}(finish, report)$). Therefore, he is fearful that he might miss the report the next morning ($\texttt{Fear}_{boss}\neg report$).

This proposition will be proved by applying Lemma 18.1: 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 18.1.

$$\mathtt{Bel}_i\, \neg \mathtt{After}_{j:\alpha}\varphi \rightarrow \mathtt{Poss}_i \mathtt{F} \neg\varphi$$

Proof.

$$\begin{aligned}
\mathtt{G}\varphi &\rightarrow \mathtt{After}_{j:\alpha}\varphi && \text{(axiom (4.7))}\\
\neg \mathtt{After}_{j:\alpha}\varphi &\rightarrow \neg \mathtt{G}\varphi && \text{(contra-position)}\\
&\rightarrow \mathtt{F}\neg\varphi && \text{(from Def}_{\mathtt{F}}\text{)}\\
\mathtt{Bel}_i\, \neg \mathtt{After}_{j:\alpha}\varphi &\rightarrow \mathtt{Bel}_i\, \mathtt{F}\neg\varphi && \text{(necessitation rule)}\\
&\rightarrow \neg \mathtt{Bel}_i\, \neg \mathtt{F}\neg\varphi && \text{(axiom D}_{\mathtt{Bel}_i}\text{)}\\
&\rightarrow \mathtt{Poss}_i \mathtt{F}\neg\varphi
\end{aligned}$$

□

Proof of Proposition 18.

$$\begin{aligned}
\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) &\rightarrow \mathtt{Goal}_i\, \varphi && \text{(from Def. 11)}\\
&\rightarrow \mathtt{Goal}_i\, \neg(\neg\varphi) && (6.6)\\
\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) &\rightarrow \mathtt{Bel}_i\, \neg \mathtt{After}_{j:\alpha}\varphi && \text{(from Def. 11)}\\
&\rightarrow \mathtt{Poss}_i \mathtt{F}(\neg\varphi) \quad \text{(from Lem. 18.1)} && (6.7)\\
\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) &\rightarrow \mathtt{Goal}_i\, \neg(\neg\varphi) \wedge \mathtt{Poss}_i \mathtt{F}(\neg\varphi)\\
& && \text{(from (6.6) and (6.7))}\\
&\rightarrow \mathtt{Fear}_i(\neg\varphi) && \text{(from Def. 2)}
\end{aligned}$$

□

6.2.2 Failure following DisTrust

A failure following the distrust of agent i on agent j on doing action α to bring about φ is the fact that agent i finally experiences that $\neg\varphi$ is true instead of φ. We assume that in this situation, agent i could feel fear-confirmed:

Proposition 19. *(Failure following DisTrust implies Fear-confirmed)*

$$\mathtt{Bel}_i\, \mathtt{X}^{-1}\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\neg\varphi \rightarrow \mathtt{FearConfirmed}_i\neg\varphi$$

Proof.

$$\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \rightarrow \mathtt{Fear}_i(\neg\varphi) \qquad \text{(from Prop. 18)}$$

$$\mathtt{Bel}_i\, \mathtt{X}^{-1}\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \rightarrow \mathtt{Bel}_i\, \mathtt{X}^{-1}\mathtt{Fear}_i(\neg\varphi)$$

$$\begin{aligned}\mathtt{Bel}_i\, \mathtt{X}^{-1}\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i(\neg\varphi) & \\ \rightarrow & \mathtt{Bel}_i\, \mathtt{X}^{-1}\mathtt{Fear}_i(\neg\varphi) \wedge \mathtt{Aware}_i(\neg\varphi) \\ \rightarrow & \mathtt{FearConfirmed}_i(\neg\varphi) \qquad \text{(from (5.3))}\end{aligned}$$

□

For instance, if the boss realizes that his assistant really did not finish the report on time ($\mathtt{Aware}_{boss}\neg report$), he feels fear-confirmed ($\mathtt{FearConfirmed}_{boss}\neg report$).

This proposition has a corollary: When we experience that what we distrusted has now happened, we feel distress about it.

Corollary 19.1.

$$\mathtt{Bel}_i\, \mathtt{X}^{-1}\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\neg\varphi \rightarrow \mathtt{Distress}_i\neg\varphi$$

Proof.

$$\begin{aligned}\mathtt{Bel}_i\, \mathtt{X}^{-1}\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\neg\varphi & \\ \rightarrow & \mathtt{FearConfirmed}_i\neg\varphi \qquad \text{(from Prop. 19)} \\ \rightarrow & \mathtt{Distress}_i\neg\varphi \qquad \text{(from Prop. 6a)}\end{aligned}$$

□

6.2.3 Success following DisTrust

A success following the distrust of agent i on agent j on doing action α to bring about φ is the fact that agent i experiences that φ is true. We assume that in this situation, agent i could feel relief:

Proposition 20. *(Success following DisTrust implies Relief)*

$$\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\varphi \rightarrow \mathtt{Relief}_i\varphi$$

Proof.

$$\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \rightarrow \mathtt{Fear}_i(\neg\varphi) \qquad \text{(from Prop. 18)}$$

$$\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \rightarrow \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Fear}_i(\neg\varphi)$$

$$\begin{aligned}\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\varphi & \\ &\rightarrow \mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{Fear}_i(\neg\varphi) \wedge \mathtt{Aware}_i\varphi \\ &\rightarrow \mathtt{Relief}_i\varphi \qquad \text{(from (5.4))}\end{aligned}$$

□

For instance, if the boss discovers that his assistant did in fact finish the report ($\mathtt{Aware}_{boss}report$), he feels relieved ($\mathtt{Relief}_{boss}report$).

This proposition has a corollary: When we experience that the trustee has done successfully the action instead of failure as we believed, we feel joyful about it.

Corollary 20.1.

$$\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\varphi \rightarrow \mathtt{Joy}_i\varphi$$

Proof.

$$\begin{aligned}\mathtt{Bel}_i\,\mathtt{X}^{-1}\mathtt{C\text{-}DisTrust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\varphi &\rightarrow \mathtt{Relief}_i\varphi \qquad \text{(from Prop. 20)} \\ &\rightarrow \mathtt{Joy}_i\neg\varphi \qquad \text{(from Prop. 6b)}\end{aligned}$$

□

Once a success follows distrust, if trustor focuses on the action's owner, trustor will feel gratitude about trustee. We accordingly formalize this relation as:

Proposition 21. *(Success following DisTrust implies Gratitude)*

$$\begin{aligned}\mathtt{Bel}_i\,\mathtt{Done}_{j:\alpha}\mathtt{I\text{-}DisTrust}_{i,j}(\alpha,\varphi) \wedge \mathtt{Aware}_i\varphi & \\ &\rightarrow \mathtt{Gratitude}_{i,j}(\alpha,\varphi)\end{aligned}$$

Proof.

$$\texttt{I-DisTrust}_{i,j}(\alpha,\varphi) \rightarrow \texttt{Goal}_i\,\varphi \wedge \texttt{Bel}_i\,(\texttt{After}_{j:\alpha}\varphi) \qquad \text{(from Def. 12)}$$

$$\begin{aligned}\texttt{Bel}_i\,\texttt{Done}_{j:\alpha}\texttt{I-DisTrust}_{i,j}(\alpha,\varphi) \wedge \texttt{Bel}_i\,\varphi \rightarrow& \\ \texttt{Bel}_i\,\texttt{Done}_{j:\alpha}(\texttt{Goal}_i\,\varphi \wedge \texttt{Bel}_i\,\texttt{After}_{j:\alpha}\varphi) \wedge \texttt{Aware}_i\varphi& \\ \rightarrow \texttt{Bel}_i\,\texttt{Done}_{j:\alpha}\texttt{Goal}_i\,\varphi \wedge \texttt{Bel}_i\,\texttt{After}_{j:\alpha}\varphi \wedge \texttt{Aware}_i\varphi& \\ \text{(from (6.5))}& \\ \rightarrow \texttt{Gratitude}_{i,j}(\alpha,\varphi) \qquad \text{(from Def. 7)}&\end{aligned}$$

□

In other term, we feel grateful about an action when we distrusted someone to do an 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 successfully the action, and that we now experience our expected outcome.

To summarize, when a success follows distrust, if the trustor focuses only on the action's outcome, s/he will feel fear-confirmed. If trustor also focuses on the action's owner, s/he will feel gratitude about trustee's action.

6.3 Discussion

6.3.1 Summary of the effect of trust/distrust on emotions

The effect of trust/distrust on emotion can be summarized in Table 6.1 in which, two first columns are a combination of the possible values of action α (*before* execution, *not done* and have *done*) and its outcome φ (*true* or *false*). Four remain columns correspond to the cases of trust or distrust. For the case of trust, we distinguish two sub-cases: trust in *action* means that the outcome φ will be achieved after execution of action α ($[\alpha]\varphi$); and trust in *inaction* means that the outcome φ will be achieved by avoiding the execution of action α ($[\sim \alpha]\varphi$). In the case of distrust, we distinguish two sub-cases: distrust based on the lack of *ability* or *intention* of trustee.

In remain cells, each emotion correspond to the case of trust or distrust of its column and the situation of action and outcome of its row. For instance, in row 2, column 3, we have $Hope\varphi$ because we have the trust and the related action is not yet executed! Cells having the symbol "x" mean that our model does not cover that corresponding case!

Hope. Hope occurs in the case that there is trust and the related action is not yet executed (or not avoided, in the case of trust in inaction - proposition 13).

Fear. Fear occurs in the case that there is distrust based on the lack of trustee's ability and the related action is not yet executed (proposition 18).

Satisfaction. Satisfaction occurs in the case that a successful outcome follows trust whatever the related action is executed (or avoided) or not (proposition 14).

Disappointment. Disappointment occurs in the case that a failed outcome follows trust whatever the related action is executed (or avoided) or not (proposition 16).

Relief. Relief occurs in the case that a successful outcome follows distrust, which is based on the lack of trustee's ability, whatever the related action is executed or not (proposition 20).

Fear-confirmed. Fear-confirmed occurs in the case that a failed outcome follows distrust, which is based on the lack of trustee's ability, whatever the related action is executed or not (proposition 19).

Anger. Anger occurs in the case that a failed outcome follows trust (betrayal): either there is a trust in action but the related action is not done, or there is a trust in inaction but the related action is done (proposition 17).

Action	Outcome	Trust in action	Trust in inaction	Distrust	
α	φ	$[\alpha]\varphi$	$[\sim \alpha]\varphi$	No ability	No intention
Before		Hopeφ	Hopeφ	Fear$\neg\varphi$	x
Not done	φ	Satisfaction	Satisfaction Gratitude	Relief	x
	$\neg\varphi$	Anger Disappointment	Disappointment	Fear-Confirmed	x
Done	φ	Satisfaction Gratitude	Satisfaction	Relief	Gratitude
	$\neg\varphi$	Disappointment	Anger Disappointment	Fear-Confirmed	x

Table 6.1: Summary of the effect of trust/distrust on emotions.

Gratitude. Gratitude occurs in the case that a success follows distrust based on the lack of trustee's intention (proposition 21) or trust (proposition 15).

In other terms, *Hope/Fear* depend on whether we trust or distrust. Confirmation-based emotions as *Satisfaction/Disappointment, Relief/Fear-confirmed* depend only on trust/distrust and theirs outcome. While compose emotions as *Anger, Gratitude* depend also on the related action's owner.

6.3.2 Exceptional cases

As showed in Table 6.1, our model does not cover all possible relations. There exists exceptional cases which are assigned a symbol "x". The reason is that we could not prove the distrust based on the lack of trustee's intention also implies the prospect of $\neg\varphi$ like this:

$\mathtt{Bel}_i\,\mathtt{After}_{j:\alpha}\varphi \wedge \mathtt{Bel}_i\,\neg\mathtt{Intend}_j\,\alpha \rightarrow \mathtt{Poss}_i\neg\varphi$

even that it seem to be quite intuitive! As a result, we could not realize the consequence cases of this relation.

6.4 Conclusion

From the formal representation of the cognitive structure of emotions, trust and distrust introduced in chapter 5, this chapter formalized the effect of trust and distrust on emotions, in the same logic introduced in chapter 4, in form of propositions. Particularly, a success following trust brings about satisfaction and/or gratitude; a failure following trust brings about disappointment and/or anger; a success following distrust brings about relief and/or gratitude; and a failure following distrust brings about fear-confirmed. These propositions will be behaviorally validated in chapter 7 (section 7.1).

Chapter 7

Behavioral Validation

Although the propositions that we proved in chapter 6 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 the concept of Trust analyzed in chapter 5; and central propositions presented in chapter 6.

Section 7.1 presents the validation of the effect of trust and distrust on emotion; Section 7.2 presents the validation of the concept of trust when we add public ground factor into its cognitive structure.

7.1 Experiment 1

7.1.1 Objective

This section reports empirical data concerning central propositions from chapter 6, related to the emotions that follow trust or distrust, paired with success or failure. More specifically, we will test whether:

- a successful outcome following trust, brings about satisfaction (Proposition 14) and/or gratitude (Proposition 15);
- a failed outcome following trust, brings about disappointment (Proposition 16) and/or anger (Proposition 17);
- a successful outcome following distrust based on lack of capacity, brings about relief (Proposition 20); and
- a successful outcome following distrust based on lack of intention, brings about gratitude (Proposition 21).

We could not test Proposition 19 for a linguistic reason: Neither in French nor in Vietnamese (the two languages used in our experiment) could we find a term equivalent to 'fear confirmed'.

7.1.2 Method

A total of 100 participants (30% French and 70% Vietnamese) took part to an online experiment; 71% were male; average age was 25 ($Min = 17$, $Max = 45$).

Participants read 8 scenarios 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).

In some case, we need to regroup intention and capacity variables into one variable (called $Trust$) in order to analyze the interaction $Trust \times Outcome$. In section 5.2.2, we defined trust as the conjunction of two beliefs (*intention* and *capacity* of the trustee) with the *public ground* between

		Capacity: Yes	Capacity: No
Intention	Yes	Trust	Distrust
	No	Distrust	Distrust

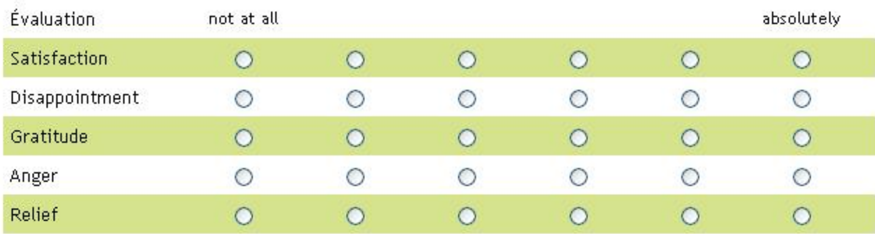
Table 7.1: Construction of the Trust variable from the capacity and intention variables

trustor and trustee. This third factor is fixed in the first experiment. We thus operationalize *Trust* as the conjunction of *Intention* and *Capacity*, and *Distrust* as the three remaining cases, as illustrated in Table 7.1.

Turning now to our materials, here is the scenario in which *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



Évaluation	not at all					absolutely
Satisfaction	○	○	○	○	○	○
Disappointment	○	○	○	○	○	○
Gratitude	○	○	○	○	○	○
Anger	○	○	○	○	○	○
Relief	○	○	○	○	○	○

Figure 7.1: The dependent variables of Experiment 1

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 our five target emotions: satisfaction, disappointment, relief, anger and gratitude. Ratings used a 6-point scale from 1 (*Not at all*) to 6 (*Absolutely*) as described in Figure 7.1.

7.1.3 Results

Language was entered as a control variable in all statistical analyses, but added only a small overall effect on participants' responses, and will not be discussed any further.

Descriptive statistics are displayed in Table 7.2. 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 × Trust, $F(1, 98) = 8.8$, $p < .01$, $\eta_p^2 = .08$, reflecting the fact that success was even more pleasant in case of trust. The

Emotion	Outcome	Trust	Distrust
Satisfaction	Success	**4.9 (1.5)**	4.6 (1.6)
	Failure	1.1 (0.5)	1.4 (1.0)
Relief	Success	2.8 (1.9)	**3.6 (1.9)**
	Failure	1.3 (1.0)	1.3 (0.9)
Disappointment	Success	1.1 (0.6)	1.3 (0.8)
	Failure	**4.6 (1.7)**	3.2 (1.4)
Anger	Success	1.1 (0.3)	1.2 (0.8)
	Failure	**3.9 (1.8)**	2.7 (1.5)
Gratitude	Success	**3.7 (1.7)**	**4.3 (1.7)**
	Failure	1.2 (0.6)	1.4 (0.9)

Table 7.2: Mean and standard deviations of affective ratings, as a function of Trust and Outcome.

data are in line with what was expected from Proposition 14.

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 $F(1, 98) = 12.3$, $p < .001$, $\eta_p^2 = .31$, reflecting the fact that Trust only had an effect when Outcome was a failure.

		Capacity	
		Yes	No
Intention	Yes	2.8	**3.5**
	No	3.7	**3.6**

Table 7.3: Mean and standard deviations of Relief ratings, as a function of Capacity and Intention (Outcome = success).

For more detail, we broke down Trust into its two components. Table 7.3 displays the effect of *Capacity* and *Intention* on Relief (in case of Outcome = success).

This analysis detected a main effect of Intention, $F(1, 79) = 14.4$, $p <$

.001, and the interaction effect $Intention \times Capacity$, $F(1,79) = 9.0$, $p < .004$, but no significant effect of Capacity, $F(1,79) = 0.4$, $p < .53$. Capacity only had an effect when there was Intention: Relief score is higher without capacity (M=3.5) than with capacity (M=2.8). This reflects our expectation (Proposition 20) that relief is especially high when a success is obtained despite of distrust based on the lack of trustee's capacity.

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, $F(1,98) = 44.7$, $p < .001$, $\eta_p^2 = .11$. This interaction reflects our expectation (Proposition 16) that disappointment is especially high when a failure is obtained despite of trust.

Anger

The analysis detected main effects of Trust, $F(1,98) = 22.4$, $p < .001$, $\eta_p^2 = .18$; and Outcome, $F(1,98) = 245$, $p < .001$, $\eta_p^2 = .71$. However, these main effects were qualified by an interaction effect Trust $\times$ Outcome, $F(1,98) = 37.5$, $p < .001$, $\eta_p^2 = .28$. This interaction reflects our expectation (Proposition 17) that anger is especially high when trust is followed by failure.

Gratitude

The analysis of variance detected a large effect of Outcome, $F(1,98) = 272$, $p < .001$, accounting for most of the observed variance, $\eta_p^2 = .73$. In other terms, Gratitude is largely predicted by Outcome alone. The analysis, however, also detected a comparatively small interaction effect involving Outcome $\times$ Trust, $F(1,98) = 3.2$, $p < .01$, $\eta_p^2 = .03$, reflecting the fact that a positive 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. Table 7.4 shows the effect of *Capacity*

		Capacity	
		Yes	No
Intention	Yes	**3.7**	3.9
	No	**4.8**	**4.4**

Table 7.4: Mean and standard deviations of Gratitude ratings, as a function of Capacity and Intention (Outcome = success).

and *Intention* on Gratitude (in case of Outcome = success).

The analysis detected a large effect of Intention, $F(1,79) = 20.5$, $p < .001$, and a small interaction effect $Intention \times Capacity$, $F(1,79) = 4.7$, $p < .03$, but no significant effect of Capacity, $F(1,79) = 0.3$, $p < .59$. Capacity only had an effect when there was Intention: Without Intention, Gratitude is high, independently from Capacity, cf. Table 7.4. This reflects our expectation that gratitude is high when a success is obtained despite of distrust based on the lack of trustee's intention (Proposition 21).

7.1.4 Discussion

Results support our formal propositions: success following trust brings about satisfaction and/or gratitude; failure following trust brings about disappointment and/or anger; success following distrust based on the lack of trustee's capacity brings about relief; and success following distrust based on the lack of trustee's intention brings about gratitude.

The success of this behavioral validation gives strong support to our approach, which is shown to capture lay users' intuitions about the effect of trust/distrust on emotions.

However, results also show other effects of trust/distrust on emotions, which are significant at the behavioral level, but that we could not, until now, capture at the formal level: success following distrust may also bring about satisfaction; success following distrust based on the lack of trustee's intention may also bring about some degree of relief; and success following distrust based on the lack of trustee's capacity may also bring about some degree of gratitude.

Moreover, at the formal level, we could predict that gratitude could come from success following trust (Proposition 15) or distrust based on the lack of trustee's intention (Proposition 21). What we could not predict, is that the intensity of gratitude would be significantly higher in the case of distrust. Our logic can not, for now, distinguish between different degrees of intensity of the same emotion. This limitation is a challenge that we must take up in the future if we want to fully account for our behavioral data.

7.2 Experiment 2

7.2.1 Objective

In section 5.2.2, we analyzed trust as the conjunction of the trustor's beliefs in the capacity and intention of the trustee, but also the *public ground* between trustor and trustee. The goal of Experiment 2 is to obtain evidence for the role of this new, public ground factor.

7.2.2 Method

A totally of 241 participants contributed to an online experiment (52% French, 48% Vietnamese, 72% were male, average age was 27).

Participants were randomly assigned to either group of a 2-level between-subject design, where Public Ground was the manipulated variable.

In all scenarios, the belief of the trustor about the *intention* and *ability* of trustee was always fixed; and the outcome was always a failure. Four scenarios were proposed to all participants:

Scenario 1

> Charles and Matthew are two neighboring farmers. Charles is a good hunter. Recently, a wolf attacked their lambs. Matthew would kill the wolf but he is not a good enough shooter.

> Charles leaves his house with a gun but Matthew does not ask him to kill the wolf. Matthew believes that Charles is able to use his gun and that he has intention to kill the wolf.
>
> The day after the return of Charles, Matthew discovers that Charles has not yet killed the wolf. In your opinion, what does he feel vis-à-vis Charles?

In the public ground version, "Charles leaves his house with a gun but Matthew does not ask him to kill the wolf" was replaced with "Charles leaves his house with a gun and Matthew asks him to kill the wolf".

Scenario 2

> Christelle and Peter are married and have a child. Christelle need to fetch the child at school, but she must attend a long meeting for work that was not provided. She wishes that Peter cares for the child.
>
> Christelle did not ask Pierre to pick up their child. Christelle believes that Peter is able to fetch the child and that he intends to do it.
>
> When she comes home late, she discovers that Peter has not yet fetched the child. In your opinion, what she does feel vis-à-vis Pierre?

In the public ground version, "Christelle did not ask Pierre to pick up their child" was replaced with "Christelle asked Pierre to pick up their child".

Scenario 3

> Bob and John are two friends. In fact, Bob was in prison because he hit a girl. He did not tell anyone his secret except John. Bob wants to hide his secret.
>
> Bob did not ask John to keep his secret. Bob believes that John is able the secret and that he intends to do it.

> One day, Bob discovers that John has spoken to someone his secret. In your opinion, what does he feel vis-à-vis John?

In the public ground version, “Bob did not ask John to keep his secret” was replaced with “Bob asked John to keep his secret”.

Scenario 4

> 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.
>
> Mr. Boss did not ask Mr. Support to prepare it and put it on his desk before tomorrow morning. Mr. Boss believes that Mr. Support has the intention and that Mr. Support is able to prepare the report in time.
>
> The morning after, Mr. Boss has no report on his desk when he arrives. In your opinion, what does he feel vis-à-vis Mr. Support?

In the public ground version, “Mr. Boss did not ask Mr. Support to prepare it and put it on his desk before tomorrow morning” was replaced with “Mr. Boss asked Mr. Support to prepare it and put it on his desk before tomorrow morning”.

Évaluation	not at all					absolutely
Satisfaction	○	○	○	○	○	○
Disappointment	○	○	○	○	○	○
Gratitude	○	○	○	○	○	○
Anger	○	○	○	○	○	○
Relief	○	○	○	○	○	○

Figure 7.2: Dependent variables in Experiment 2

After reading each scenario, participants rated the extent to which the main character would feel each of 5 our target emotions: satisfaction, disappointment, relief, anger and gratitude as described in Figure 7.2. The 6-point scale was recoded from -2.5 to +2.5.

7.2.3 Results

Language was entered as a control variable in all statistical analyses, but added only a small overall effect on participants' responses, and will not be discussed any further.

Overall, and in line with the results of Experiment 1, failure in was judged to trigger negative emotions such as *Anger* and *Disappointment*, as showed in Table 7.5. However, *public ground* had a main effect in the case of *Anger*: $F = 61.4$, $p < .001$.

	Public ground			
	Yes	No	F	p
Satisfaction	-2.3 (0.4)	-2.1 (0.7)	9.1	.003
Disappointment	1.5 (1.1)	1.0 (1.2)	9.6	.002
Relief	-2.0 (0.8)	-2.0 (0.8)	0.1	.728
Gratitude	-2.1 (0.6)	-2.1 (0.7)	0.1	.764
Anger	**1.0 (1.2)**	**-0.3 (1.3)**	**61.4**	**.000**

Table 7.5: Mean and standard deviations of affective ratings, as a function of Public Ground.

Anger's score is significantly higher with public ground ($M = 1.0$, $SD < 1.2$), and there is no anger in the absence of public ground ($M = -0.3$, $SD < 1.3$). This is in line with our analysis, presented in section 5.2.2.

7.2.4 Discussion

In line with our previous results and with the literature (de Sousa [56, 236]), we assume that the absence of anger reflects the absence of betrayal, and thus the absence of previous trust. Results show that anger is absent in the absence of Public Ground. Thus we conclude that trust requires

Public Ground. Our argumentation that public ground factor is a necessary component of trust is now successfully validated at the behavioral level.

Regarding the score of disappointment, there is no qualitative difference between two groups: disappointment is always present in both cases. However, its intensity is a bit higher in the case of having trust. In the first group, disappointment may appear because the case of trustee's intention and capacity, without public ground would correspond to the concept of reliance, which is well distinguished from the concept of trust by Baier [12]. Intuitively, a failure following reliance could bring about disappointment.

Public ground also has an effect in the case of Satisfaction in spite there is no qualitative difference between its score in two groups. Satisfaction is always absent in both cases because of failed outcome. This is in line with the results of Experiment 1.

7.3 Conclusion

This chapter presented the validation of our formal propositions and concepts, using the methods of experimental psychology.

Results of the first experiment wholly supported our central propositions about the effect of trust/distrust on emotions: a successful outcome following trust brings about satisfaction and/or gratitude, a failed outcome following trust brings about disappointment and/or anger, a successful outcome following distrust based on the lack of trustee's capacity brings about relief, and a successful outcome following distrust based on the lack of trustee's intention brings about gratitude.

Results of the second experiment supported our claim that the public ground between trustor and trustee is a necessary component in the cognitive structure of trust.

The success of these behavioral validations gives strong support to our approach, which is shown to capture lay users' intuitions about the effect of trust/distrust on emotions.

Conclusion

Treatment of emotion is not an easy mission, specially making up to the computer the ability of treatment of emotion because: (i) emotion itself is not easy to understand, it is a subjective attitude mental which depends not only upon the real situation but also depends upon how does the owner feel; (ii) there is not yet an agreement among researches on the concept of emotion, it simultaneously involves several types of interdependent reactions: cognitive, physiological, biological and cultural, etc. that relate to different fields of research, so most definitions of emotions are partial since they only refer to one of their aspects; (iii) the computer can not feel emotion like the way that human does, it can possibly recognize emotions based on their cognitive factors, but what are exactly the cognitive factors for each emotion is not easy to determine; and (iv) even if we have a good foundation model for emotion, how can we make the computer treating (to recognize, understand and reasoning) emotion in a natural human-like way is not yet easy one.

Despite many difficulties in assigning to the computer the human-like abilities of understand, represent and reason about emotions, we constructed a logical framework to formally represent trust/distrust and emotions in a formal language. This logic enables us to lay bare the formal relations between trust and emotions, particularly the effect of trust and distrust on emotions. The logic we offer is a combination of the logic of beliefs and choices, and the logic of time.

The main contributions of this thesis can be summarized in three facts.

Firstly, we proposed a logical framework which enables to represent the cognitive structure of some concept of emotions, trust and distrust in a formal language. This representation contributes in improving precedent approaches. Regarding emotion, our formal definition respects well the instant nature of emotion by considering a time factor in its decompositions. On the trust's concept, our definition capitalizes the fact of trustor's belief about the public ground between him and trustee.

Secondly, we also formalized the effect of trust and distrust on some emotions in the same logic by considering potential possibilities relating to the outcome of trust and distrust. Its objective is to predict what is the next potential emotion based on the user's current attitude mental, particularly in the case of success or failure following trust and distrust: a success following trust brings about satisfaction and/or gratitude; a failure following trust brings about disappointment and/or anger; a success following distrust brings about relief and/or gratitude; and a failure following distrust brings about fear-confirmed. The formal representation of these effects enables computers have the ability to reason more naturally about human emotion based upon some causal attitude which has a tight and natural relation with emotions: trust and distrust mental states.

Thirdly, these formalized effects of trust and distrust on emotions are also behaviorally validated by collecting the empirical data. The support of empirical data to our formal propositions contributes to the psychological analysis, at the behavioral level, the effect of trust and distrust on emotions, particularly on satisfaction, gratitude, disappointment and anger. Additionally, the success of our validation on the concept of trust also behaviorally confirms the fact that it is necessary to have the public ground factor in the decompositions of trust's cognitive structure.

However, our work remains some limitation points which could be improved in the future. First, the intensity of emotions. Although the added time factor helps us to respect the instant nature of emotion, it is not enough utility to help us to cover the intensity of emotions. Our formal emotions exist in boolean values and during the time it presents, its intensity is no change until the emotion disappears. Second, we are currently on the *appraisal* step, the first step on the human emotional process. The remaining step, *coping* which leads to a subsequent adaptive modification of behavior, is not yet considered in out framework. These limitations also are possibly considered as potential fields to our work in the future.

Bibliography

[1] A. Abdul-Rahman and S. Hailes. Supporting trust in virtual communities. In *Proceedings of the Hawaii's International Conference on Systems Sciences*, Maui, Hawaii, 2000.

[2] Carole Adam. *Emotions: from psychological theories to logical formalization and implementation in a BDI logic*. PhD thesis, University of Paul Sabatier, Toulouse, France, July 2007.

[3] Carole Adam, Andreas Herzig, and Dominique Longin. A logical formalization of the OCC theory of emotions. *Synthese*, 168(2):201–248, 2009.

[4] C. Adams and S. Farrell. RFC2510 - internet X.509 public key infrastructure certificate management protocols. Technical report, http://www.cis.ohio-state.edu/htbin/rfc/rfc2510.html, 1999.

[5] J. K. Aggarwal and Q. Cai. Human motion analysis: A review. *Computer Vision and Image Understanding*, 73:428–440, 1999.

[6] Philippe Aghion, Yann Algan, Pierre Cahuc, and Andrei Shleifer. Regulation and distrust. *SUS.DIV–CEPR–PSEConference of Models of Cultural Dynamics and Diversity*, 2008.

[7] Amazon. Amazon auctions. http://auctions.amazon.com, 2010.

[8] Poul Houman Andersen and Rajesh Kumar. Emotions, trust and relationship development in business relationships: A conceptual model for buyer-seller dyads. *Industrial Marketing Management*, 35:522–535, 2006.

[9] Magda Arnold. *Emotion and Personality*. Columbia University Press, New York, 1960.

[10] Donovan Artz and Yolanda Gil. A survey of trust in computer science and the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2007.

[11] Stylianos Asteriadis, Paraskevi Tzouveli, Kostas Karpouzis, and Stefanos Kollias. Estimation of behavioral user state based on eye gaze and head pose–application in an e-learning environment. *Multimedia Tools Appl.*, 41(3):469–493, 2009.

[12] Annette Baier. Trust and antitrust. *Journal of Ethics*, 96(2):231–260, 1986.

[13] Jack Barbalet. Trust and uncertainty: The emotional basis of rationality. University of Leicester, 2008.

[14] Monica Y Bartlett and David DeSteno. Gratitude and prosocial behavior: helping when it costs you. *Psychol Sci*, 17(4):319–325, 2006.

[15] Joseph Bates. The role of emotion in believable agents. *Communications of the ACM*, 37:122–125, July 1994.

[16] L. C. Becker. Trust as noncognitive security about motives. *Ethics*, 107(1):43–61, 1996.

[17] Christopher J. Beedie, Peter C. Terry, and Andrew M. Lane. Distinctions between emotion and mood. *Cognition and emotion*, 19(6):847–878, 2005.

[18] Nuel D. Belnap, Jr., Michael Perloff, and Ming Xu. *Facing the Future: Agents and Choices in Our Indeterminist World*. Oxford University Press, Oxford, 2001.

[19] J. Bigün, E. S. Bigun, B. Duc, and S. Fischer. Expert conciliation for multi modal person authentication systems by bayesian statistics. In *Proceedings of the International Conference on Audio and Video-Based Biometric Person Authentication (AVBPA)*, pages 291–300, Crans-Montana, Switzerland, 1997.

[20] M. Black and Y. Yacoob. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In *Proceedings of the International Conference on Computer Vision*, pages 374–381, 1995.

[21] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Scie.* Cambridge University Press, Cambridge, 2001.

[22] M. Blaze. RFC2704 - the KeyNote trust management system (version 2). Technical report, http://www.crypto.com/papers/rfc2704.txt, 1999.

[23] M. Blaze. Using the KeyNote trust management system. Technical report, http://www.crypto.com/trustmgt/kn.html, 1999.

[24] M. Blaze, J. Feigenbaum, and J. Lacy. Decentralized trust management. In *Proceedings of the IEEE Conf. Security and Privacy*, Oakland, California, USA, 1996.

[25] Jean-François Bonnefon, Dominique Longin, and Manh Hung Nguyen. A Logical Framework for Trust-Related Emotions. *Electronic Communications of the EASST, Formal Methods for Interactive Systems 2009*, 22:(online), 2009.

[26] Jean-François Bonnefon, Dominique Longin, and Manh Hung Nguyen. Relation of Trust and Social Emotions: A Logical Approach. In *IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT), Milano, Italy, 15/09/09-18/09/09*, pages 289–292. IEEE, 2009.

[27] M. E. Bratman, D. Israel, and M. E. Pollack. Plans and resource bounded practical reasoning. *Computational Intelligence*, 4:349–355, 1988.

[28] Michael Bratman. *Intentions, Plans, and Practical Reason.* Harvard University Press, 1987.

[29] J.W. Brehm. The intensity of emotion. *Personality and Social Psychology Review*, 3:2–22, 1999.

[30] C. D. Broad. *Emotion and Sentiment.* Allen & Unwin, London, 1971.

[31] Roberto Brunelli and Daniele Falavigna. Person identification using multiple cues. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(10):955–966, 1995.

[32] Herrington 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.

[33] M. Burrows, M. Abadi, and R. M. Needham. A logic of authentication. *ACM Trans. Comp. Systems*, 8(1):18–36, 1990.

[34] Nick Campbell. Perception of affect in speech - towards an automatic processing of paralinguistic information in spoken conversation. In *Proceedings of the 8th International Conference on Spoken Language Processing (INTERSPEECH 2004 - ICSLP)*, Jeju Island, Korea, October 2004.

[35] Nick Campbell. Developments in corpus-based speech synthesis: Approaching natural conversational speech. *IEICE - Trans. Inf. Syst.*, E88-D(3):376–383, 2005.

[36] W. B. Cannon. The james-lange theory of emotion: A critical examination and an alternative theory. *American Journal of Psychology*, 39:10–124, 1927.

[37] Walter Cannon. *Bodily Changes in Pain, Hunger, Fear and Rage*. Appleton, New York, 2nd edition, 1929.

[38] J. Carter, E. Bitting, and A. Ghorbani. Reputation formalization for an information-sharing multi-agent sytem. *Computational Intelligence*, 18(2):515–534, 2002.

[39] C. Castelfranchi, R. Falcone, and E. Lorini. A non-reductionist approach to trust. In J. Goldbeck, editor, *Computing with Social Trust*, pages 45–72. Springer, Berlin, 2008.

[40] Cristiano Castelfranchi and Rino Falcone. Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. In *Proceedings of the International Conference on Multi-Agent Systems (ICMAS'98)*, pages 72–79, Paris, France, 1998.

[41] Cristiano Castelfranchi and Rino Falcone. *Trust Theory: A socio-cognitive and computational model*. Wiley Series in Agent Technology. Wiley, 2010.

[42] Brian Chellas. *Modal Logic - An Introduction*. Cambrige University Press, Cambrige, 1980.

[43] L. Chen. *Joint processing of audio-visual information for the recognition of emotional expressions in human-computer interaction*. PhD thesis, University of Illinois at Urbana-Champaign, 2000.

[44] C. Chiu, Y. Chang, and Y. Lai. The analysis and recognition of human vocal emotions. In *Proceedings of the International Computer Symposium*, pages 83–88, 1994.

[45] Y.-H. Chu. REFEREE: Trust management for web applications. Technical report, http://www.research.att.com/ jf/pubs/www6-97.html, 1997.

[46] Ze-Jing Chuang and Chung-Hsien Wu. Emotion recognition from textual input using an emotional semantic network. In *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP-2002)*, pages 2033–2036, Denver, Colorado, USA, September 2002.

[47] 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*, 8(2):244–263, 1986.

[48] A. M. Coddington and M. Luck. Towards motivation-based plan evaluation. In I. Russell and S. Haller, editors, *Recent Advances in Artificial Intelligence: Proceedings of the Sixteenth International FLAIRS Conference*, pages 298–302. AAAI Press, 2003.

[49] I. Cohen, N. Sebe, A. Garg, L. Chen, and T. Huang. Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision and Image Understanding*, 91(1–2):160–187, 2003.

[50] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.

[51] E. Cosatto, H.P. Graf, G. Potamianos, and J. Schroeter. Audio-visual selection process for the synthesis of photo-realistic talking-head animations. In *IEEE International Conference on Multimedia and Expo, ICME 2000*, volume 4, pages 619–622, 2003.

[52] J.B. Cullen, J.L. Johnson, and T. Sakano. Success through commitment and trust: The soft side of strategic alliance management. *Journal of World Business*, 35(3):223–240, 2000.

[53] Antonio R. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain.* Putnam Pub Group, 1994.

[54] Charles Darwin. *The Expression of the Emotions in Man and Animals.* Dover Publications, Inc., second edition, 2007.

[55] R.J. Davidson. On emotion, mood and related affective constructs. In P. Ekman and R.J. Davidson, editors, *The nature of emotion*, pages 51–55. Oxford University Press, Oxford, England, 1994.

[56] Ronald de Sousa. *The Rationality of Emotion.* MIT Press, 6 edition, 2001.

[57] Frank Dellaert, T. Polzin, and A. Waibel. Recognizing emotion in speech. In *Proceedings of the ICSLP '96, Philadelphia, PA*, pages 1970–1973, October 1996.

[58] Robert Demolombe and Emiliano Lorini. Trust and norms in the context of computer security: toward a logical formalization. In R. Van der Meyden and L. Van der Torre, editors, *International Workshop on Deontic Logic in Computer Science (DEON), Luxembourg, 15/07/2008-18/07/2008*, volume 5076 of *LNCS*, pages 50–64. Springer-Verlag, 2008.

[59] M. Deutsch. *The resolution of conflict: Constructive and destructive processes.* New Haven, CN: Yale University Press, 1973.

[60] DLNA. DLNA: Enjoy your music, photos and videos, anywhere anytime. http://www.dlna.org/en/consumer/home, 2006.

[61] Sidney D'Mello, Rosalind W. Picard, and Arthur Graesser. Toward an affect-sensitive autotutor. *IEEE Intelligent Systems*, 22(4):53–61, 2007.

[62] Jennifer R. Dunn and Maurice E. Schweitzer. Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology*, 88(5):736–748, 2005.

[63] Cahn J. E. The generation of affect in synthesized speech. *Journal of the American Voice I/O Society*, pages 1–19, 8, July 1990.

[64] eBay. ebay. http://www.ebay.com, 2010.

[65] E. Eide, A. Aaron, R. Bakis, W. Hamza, M. Picheny, and J. Pitrelli. A corpus-based approach to <ahem/> expressive speech synthesis. In *in 5th ISCA Workshop on Speech Synthesis*, pages 79–84, 2004.

[66] P. Ekman. Moods, emotions and traits. In P. Ekman and R.J. Davidson, editors, *The nature of emotion*, pages 56–58. Oxford University Press, Oxford, England, 1994.

[67] P. Ekman. Strong evidence for universals in facial expressions: A reply to russellś mistaken critique. *Psychological Bulletin*, 115(2):268–287, 1994.

[68] P. Ekman and W. Friesen. *Facial Action Coding System: Investigator's Guide*. Consulting Psychologists Press, 1978.

[69] Jon Elster. *Alchemies of the Mind: Rationality and the Emotions*. Cambridge University Press, 1999.

[70] E. Allen Emerson. Temporal and modal logic. In *Handbook of Theoretical Computer Science, Volume B: Formal Models and Sematics (B)*, pages 995–1072. 1990.

[71] E. Allen Emerson, Michael Evangelist, and Jai Srinivasan. On the limits of efficient temporal decidability (extended abstract). In *Proceedings, Fifth Annual IEEE Symposium on Logic in Computer Science, 4-7 June 1990, Philadelphia, Pennsylvania, USAS*, pages 464–475. IEEE Computer Society, 1990.

[72] B. Esfandiari and S. Chandrasekharan. On how agents make friends: Mechanisms for trust acquisition. In *Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies*, pages 27–34, Montreal, Canada, 2001.

[73] I. Essa and A. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):757–763, 1997.

[74] N.L. Etcoff and J.J. Magee. Categorical perception of facial expressions. *Cognition*, 44:227–240, 1992.

[75] Rino Falcone and Cristiano Castelfranci. Social trust: A cognitive approach. In Cristiano Castelfranchi and Yao-Hua Tan, editors, *Trust and Deception in Virtual Societibes*, pages 55–90. Kluwer Academic Publishers, 2001.

[76] B. Fehr, M. Baldwin, L. Collins, S. Patterson, and R. Benditt. Anger in close relationships: an interpersonal script analysis. *Personality and Social Psychology Bulletin*, 25:299–312, 1999.

[77] FIPA. Abstract architecture specification, version j. Technical report, http://www.fipa.org/repository, 2002.

[78] M.J. Fischer and R.E. Ladner. Propositional dynamic logic of regular programs. *Journal of Computer and System Sciences*, 18(2):194–211, 1979.

[79] Joseph P. Forgas. Mood and judgment: The affect infusion model *aim*. *Psychological Bulletin*, 117(1):39–66, 1995.

[80] N.H. Frijda. Varieties of affect: Emotions and episodes. In P. Ekman and R.J. Davidson, editors, *The nature of emotion*, pages 59–67. Oxford University Press, Oxford, England, 1994.

[81] Nico H. Frijda. *The Emotions: Studies in Emotion and Social Interaction.* Edition de la Maison des Sciences de l'Homme. Cambridge University Press, Paris, 1986.

[82] Robert W. Frischholz and Ulrich Dieckmann. Bioid: A multimodal biometric identification system. *Computer*, 33(2):64–68, 2000.

[83] D. Gambetta. Can we trust trust? In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations*, pages 213–237. Basil Blackwell, New York, 1990.

[84] D. Garlan, D. P. Siewiorek, A. Smailagic, and P. Steenkiste. Project Aura: Toward distraction-free pervasive computing. *IEEE Pervasive Computing*, 1:22–31, 2002.

[85] Benoit Gaudou, Andreas Herzig, and Dominique Longin. A Logical Framework for Grounding-based Dialogue Analysis. In Wiebe van der Hoek,

Alessio Lomuscio, Erik de Vink, and Mike Wooldridge, editors, *International Workshop on Logic and Communication in Multi-Agent Systems (LCMAS), Edinburgh, Scotland, UK, 01/08/2005*, volume 157 of *Electronic Notes in Theoretical Computer Science (ENTCS)*, pages 117–137, http://www.elsevier.com/, 2006. Elsevier.

[86] Michael Georgeff, Barney Pell, Martha Pollack, Milind Tambe, and MichaelWooldridge. The belief-desire-intention model of agency. *In J.P. Muller et al. (Eds.): ATAL'98, LNAI 1555*, pages 1–10, 1999.

[87] Carlos Gershenson. Modelling emotions with multidimensional logic. In *NAFIPS'99*. IEEE, 1999.

[88] M. E. Glickman. Parameter estimation in large dynamic paired comparison experiments. *Applied Statistics*, 48:377–394, 1999.

[89] Paul Gochet and Pascal Gribomont. Epistemic Logic. In Dov Gabbay and John Woods, editors, *Twentieth Century Modalities*, volume 7 of *Handbook of the History of Logic*, pages 99–195. Elsevier, amsterdam edition, 2006.

[90] T. Govier. *Social Trust and Human Communities*. Montreal and Kingston: McGill-Queenś University Press, 1997.

[91] Arthur Graesser, Patrick Chipman, Brandon King, Bethany McDaniel, and Sidney DḾello. Emotions and learning with AutoTutor. In *Proceeding of the 2007 conference on Artificial Intelligence in Education*, pages 569–571, Amsterdam, The Netherlands, The Netherlands, 2007. IOS Press.

[92] Tyrone Grandison and Han Reichgelt. Formalizing trust-based decision making in electronic commerce transactions. In *Proceedings of the First International Workshop on Context Awareness and Trust (CAT07)*, pages 13–28, Moncton, New Brunswick, Canada, July 2007.

[93] Tyrone Grandison and Morris Sloman. A survey of trust in internet application. *IEEE, Communications Surveys, Fourth Quarter*, 2000.

[94] Tyrone Grandison and Morris Sloman. Specifying and analysing trust for internet applications. In *Proceedings of the 2nd IFIP Conference on e-Commerce, e-Business, e-Government*, Lisbon, Portugal, October 2002.

[95] J. Gratch and S. Marsella. A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research*, 5(4):269–306, 2004.

[96] Jonathan Gratch and Stacy Marsella. Lessons from emotion psychology for the design of lifelike characters. *Journal of Applied Artificial Intelligence (special issue on Educational Agents - Beyond Virtual Tutors)*, 19(3–4):215–233, 2005.

[97] R. Grimm. One.world: Experiences with a pervasive computing architecture. *IEEE Pervasive Computing*, 3:22–30, 2004.

[98] Joseph Y. Halpern and Yoram Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54(3):319–379, 1992.

[99] R. Hardin. *Trust and Trustworthiness.* NY: Russell Sage Foundation, New York, 2002.

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

[101] D. O. Hebb. *The organization of behavior.* Wiley, New York, 1949.

[102] A. Herzberg, Y. Mass, J. Michaeli, D. Naor, and Y. Ravid. Access control meets public key infrastructure, or: Assigning roles to strangers. In *Proceedings of the IEEE Symposium on Security and Privacy, Oakland*, pages 2–14. IEEE Computer Society, 2000.

[103] 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.

[104] Andrea Herzig and Emiliano Lorini. A logic of trust and reputation. *Logic Journal of the IGPL*, 2009.

[105] Andreas Herzig and Dominique Longin. C&L intention revisited. In *Proceedings of Int. Conf. of knowledge representation and reasoning KR'04*, pages 527–535. Morgan Kaufmann, 2004.

[106] D. K. J. Heylen, A. Nijholt, H. J. A. op den Akker, and M. Vissers. Socially intelligent tutor agents. In R. Aylett, D. Ballin, and T. Rist, editors, *Proceedings Intelligent Virtual Agents (IVA 2003), Kloster Irsee, Germany*, volume 2792 of *Lecture Notes in Artificial Intelligence*, pages 341–347, Berlin, September 2003. Springer Verlag.

[107] J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions.* Cornell University Press, Ithaca, 1962.

[108] Jaakko Hintikka. *The Logic of Epistemology and the Epistemology of Logic.* Kluwer - The language of science, 1989.

[109] T.K. Ho, J.J. Hull, and S.N. Srihari. Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:66–75, 1994.

[110] J. G. Holmes. Trust and the appraisal process in close relationships. In W. H. Jones and D. Perlman, editors, *Advances in personal relationships*, pages 57–104. Jessica Kingsley, London, 1991.

[111] Jun Hu, Chun Guan, Maoguang Wang, and Fen Lin. Model of emotional agent. *Z. Shi and R. Sadananda (Eds.): PRIMA 2006, LNAI 4088*, pages 534–539, 2006.

[112] A. J. Hunt and A. W. Black. Unit selection in a concatenative speech synthesis system using a large speech database. In *ICASSP 96́: Proceedings of the Acoustics, Speech, and Signal Processing, 1996. on Conference Proceedings., 1996 IEEE International Conference*, pages 373–376, Washington, DC, USA, 1996. IEEE Computer Society.

[113] Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. Developing an integrated trust and reputation model for open multi-agent systems. In *Proceedings of the 7th Int Workshop on Trust in Agent Societies*, pages 65–74, New York, USA, 2004.

[114] Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 13(2):119–154, 2006.

[115] IST-Ozone. OZONE: New technologies and services for emerging nomadic societies. http://www.hitechprojects.com/euprojects/ozone/, 2004.

[116] ITEA-AMBIENCE. Contextaware environments for ambient services. http://www.iteaoffice.org/public/project_leaflets/AMBIENCE_results_oct04.pdf, 2004.

[117] Anil K. Jain and Arun Abraham Ross. Learning user-specific parameters in a multibiometric system. In *ICIP (1)*, pages 57–60, 2002.

[118] S. S. Jajodia, P. Samarati, and V. Subrahmanian. A logical language for expressing authorizations. *Security and Information Privacy*, 1997.

[119] William James. What is an emotion? *Mind*, 9:188–205, 1884.

[120] N.R. Jennings. Commitments and conventions: The foundation of coordination in multi-agent systems. *The Knowledge Engineering Review*, 3:223–250, 1993.

[121] Andrew J.I. Jones. On the concept of trust. *Decision Support Systems*, 33:225–232, 2002.

[122] Christian Jones and Andrew Deeming. *Affective Human-Robotic Interaction*, pages 175–185. Springer-Verlag, Berlin, Heidelberg, 2008.

[123] J. I. A. Jones and B. S. Firozabadi. On the characterisation of a trusting agent - aspects of a formal approach. *Wksp. Deception, Trust and Fraud in Agent Societies*, 2000.

[124] K. Jones. Second-hand moral knowledge. *Journal of Philosophy*, 96(2):55–78, 1999.

[125] A. Jøsang. The right type of trust for distributed systems. In C. Meadows, editor, *Proc. of the 1996 New Security Paradigms Workshop*. ACM, 1996.

[126] Audun Jøsang, Claudia Keser, and Theo Dimitrakos. Can we manage trust? In *Proceedings of the 3rd International Conference on Trust Management, (iTrust)*, Paris, 2005.

[127] Stig Kanger. Law and logic. *Theoria*, 38:105–132, 1972.

[128] George Katsionis and Maria Virvou. A cognitive theory for affective user modelling in a virtual reality educational game. In *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics: The Hague, Netherlands, 10-13 October 2004*, pages 1209–1213. IEEE, 2004.

[129] Josef Kittler, Mohamad Hatef, Robert P.W. Duin, and Jiri Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:226–239, 1998.

[130] Ajay Kumar, David C. M. Wong, Helen C. Shen, and Anil K. Jain. Personal verification using palmprint and hand geometry biometric. In *Audio-and Video-Based Biometrie Person Authentication, 4th International Conference, AVBPA 2003, Guildford, UK, June 9-11, 2003 Proceedings*, pages 668–678, 2003.

[131] Bernd Lahno. On the emotional character of trust. *Journal of Ethical Theory and Moral Practice*, 4:171–189, 2001.

[132] P.J. Lang. What are the data of emotion? In V. Hamilton, G.H. Bower, and N.H. Frijda, editors, *Cognitive perspectives on emotion and motivation*, pages 173–191. Kluwer Academic Publishers, Dordrecht, Holland, 1988.

[133] A. Lanitis, C. Taylor, and T. Cootes. A unified approach to coding and interpreting face images. In *Proceedings of the International Conference on Computer Vision*, pages 368–373, 1995.

[134] Y. Lashkari, M. Metral, and P. Maes. Collaborative interface agents. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*. AAAIPress, 1994.

[135] Richard S. Lazarus. *Emotion and Adaptation.* Oxford University Press, 1991.

[136] R.S. Lazarus. The stable and the unstable in emotion. In P. Ekman and R.J. Davidson, editors, *The nature of emotion*, pages 79–85. Oxford University Press, Oxford, England, 1994.

[137] C. Lee, S.S. Narayanan, and R. Pieraccini. Recognition of negative emotions from the speech signal. In *Automatic Speech Recognition and Understanding*, 2002.

[138] Lester A. Lefton. *Psychology.* A Division of Paramount Publishing, 1994.

[139] R.J. Lewicki and C. Wiethoff. Trust, trust development, and trust repair. In M. Deutsch and P.T. Coleman, editors, *The Handbook of Conflict Resolution: Theory and Practice*, pages 86–107. Jossey-Bass, San Francisco, CA, 2000.

[140] Roy J. Lewicki, Daniel J. Mcallister, and Robert J. Bies. Trust and distrust: New relationships and realities. *The Academy of Management Review*, 23(3):384–458, 1998.

[141] Roy J. Lewicki and Edward C. Tomlinson. Trust and trust building. In Guy Burgess and Heidi Burgess, editors, *Beyond Intractability.* Conflict Research Consortium, University of Colorado, Boulder, 2003.

[142] J. D. Lewis and A. J. Weigert. Social atomism, holism, and trust. *The Sociological Quarterly*, 216(4):455–471, 1985.

[143] Lars Lindahl. *Position and Change - A study in Law and Logic.* Synthese Library 112. D. Reidel Publishing Company, Dordrecht, 1977.

[144] Christine Lisetti and Cynthia LeRouge. Affective computing in tele-home health. In *HICSS0́4: Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS0́4) - Track 6*, page 60148.1, Washington, DC, USA, 2004. IEEE Computer Society.

[145] Dominique Longin, Emiliano Lorini, and Manh Hung Nguyen. Delegation as a communicative act: a logical analysis. In *The Eighth IJCAI Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems (COIN@IJCAI'09), Pasadena, CA, USA, 11/07/09-13/07/09*, 2009.

[146] Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi, and Jiro Gyoba. Coding facial expressions with Gabor Wavelets. In *3rd International Conference on Face & Gesture Recognition (FG '98), April 14-16, 1998, Nara, Japan*, pages 200–205. IEEE Computer Society, 1998.

[147] William Lyons. *Emotion*. Cambridge University Press, Cambridge, 1980.

[148] B. MacNamee and P. Cunningham. Creating socially interactive no-player characters: The m-SIc system. *International Journal of Intelligent Games and Simulation*, 2(1):28–35, 2003.

[149] D. W. Manchala. E-commerce trust metrics and models. *IEEE Internet Comp.*, pages 36–44, 2000.

[150] G. Mandler. The nature of emotion. In J. Miller, editor, *States of mind: Conversations with psychological investigators*. BBC, London, 1983.

[151] Joel Marks. A theory of emotion. *Philosophical Studies*, 42:227–242, 1982.

[152] S. Marsella and J. Gratch. Modeling the interplay of emotions and plans in multi-agent simulations. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, Edinburgh, Scotland, 2001.

[153] S. Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, Department of Mathematics and Computer Science, University of Stirling, 1994.

[154] Stephen P. Marsh and Mark R. Dibben. *Trust, Untrust, Distrust and Mistrust - An Exploration of the Dark(er) Side*, volume 3477 of *Lecture Notes in Computer Science*, pages 17–33. Springer Berlin / Heidelberg, May 2005.

[155] Jean-Claude Martin, Radoslaw Niewiadomski, Laurence Devillers, Stéphanie Buisine, and Catherine Pelachaud. Multimodal complex emotions: Gesture expressivity and blended facial expressions. *I. J. Humanoid Robotics*, 3(3):269–91, 2006.

[156] K. Mase. Recognition of facial expression from optical flow. *IEICE Trans.*, 74(10):3474–3483, 1991.

[157] Y. Mass and O. Shehory. Distributed trust in open multi-agent systems. In R. Falcone, M. Singh, and Y. H. Tan, editors, *Trust in Cyber-societies*, pages 159–173. Springer-Verlag, Berlin, 2001.

[158] Fabio Massacci. Single step tableaux for modal logics. *Journal of Automated Reasoning*, 24(3):319–364, 2000.

[159] D.J. McAllister. Affect and cognition-based trust as foundation for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1):24–59, 1995.

[160] L. Thorne McCarty. Permissions and obligations. *IJCAI-83*, 1983.

[161] M. E. McCullough, S. D. Kilpatrick, R. A. Emmons, and D. B. Larson. Is gratitude a moral affect? *Psychological Bulletin*, 127:249–266, 2001.

[162] D. H. McKnight and N. L. Chervany. The meanings of trust. Technical report, 395 Hubert H. Humphrey Center, 271-19th Avenue South, Minneapolis, MN 55455, 1996.

[163] C. McLeod. *Self-Trust and Reproductive Autonomy*. Cambridge, MA: MIT Press, 2002.

[164] John-Jules Ch. Meyer. Reasoning about emotional agents. *International Journal of Intelligent Systems*, 21(6):601–619, 2006.

[165] Maria Miceli and Cristiano Castelfranchi. Hope: The power of wish and possibility. *Theory & Psychology*, 20(2):251–276, 2010.

[166] J. Miller, P. Resnick, and D. Singer. PICS rating services and rating systems (and their machine readable descriptions) version 1.1. Technical report, http://www.w3.org/TR/REC-PICS-services.

[167] W.N. Morris. A functional analysis of the role of mood in affective systems. In M. S. Clarke, editor, *Emotion*, pages 257–293. Sage, Newbury Park, CA, 1992.

[168] David Moshman, John A. Glover, and Roger H. Bruning. *Developemental Psychlogy*. Litle, Brown & Company, Canada, 1987.

[169] Sylvie J.L. Mozziconacci and Dik J. Hermes. Expression of emotion and attitude through temporal speech variations. In *in Proc. 2000 Int. Conf. Spoken Language Processing (ICSLP 2000)*, pages 1–24. Birkhauser, 2000.

[170] Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. Notions of reputation in multi-agent systems: A review. In *Proceedings of the First International Conference on Autonomous Agents and MAS*, pages 280–287, Bologna, Italy, July 2002. ACM.

[171] I. Murray and J. Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature of human vocal emotion. *Journal of the Acoustic Society of America*, 93(2):1097–1108, 1993.

[172] I. Murray and J. Arnott. Synthesizing emotions in speech: Is it time to get excited? In *Proceedings of the International Conference on Spoken Language Processing*, pages 1816–1819, 1996.

[173] Samia Nefti, Farid Meziane, and Khairudin Kasiran. A fuzzy trust model for e-commerce. In *Proceedings of the Seventh IEEE International Conference on E-Commerce Technology (CEC0́5)*, pages 401–404, 2005.

[174] Network Associates Inc. *An Introduction to Cryptography, in PGP 6.5.1 User's Guide.*

[175] Jerome Neu. *A Tear is an Intellectual Thing: the Meaning of Emotions.* Oxford University Press, Oxford, New York, 2000.

[176] Radoslaw Niewiadomski, Elisabetta Bevacqua, Maurizio Mancini, and Catherine Pelachaud. Greta: an interactive expressive ECA system. In Carles Sierra, Cristiano Castelfranchi, Keith S. Decker, and Jaime Simão Sichman, editors, *8th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009), Budapest, Hungary, May 10-15, 2009, Volume 2*, pages 1399–1400. IFAAMAS, 2009.

[177] Martha Nussbaum. *Upheavals of Thought: The Intelligence of Emotions.* Cambridge University Press, Cambridge, 2001.

[178] Justin Oakley. *Morality and the Emotions.* Routledge and Kegan Paul, London, 1992.

[179] K. Oatley and J.M. Jenkins. Human emotions: Function and dysfunction. *Annual Review of Psychology*, 43:55–85, 1992.

[180] Keith Oatley. The sentiments and beliefs of distributed cognition. In Nico H. Frijda, Antony S. R. Manstead, and Sacha Bem, editors, *Emotions and beliefs: How Feelings Influence Thoughts*, chapter 4, pages 78–107. Cambridge University Press, The edinburgh Building, Cambridge CB2 2RU, UK, 2000.

[181] Magalie Ochs, Karl Devooght, David Sadek, and Catherine Pelachaud. A computational model of capability-based emotion elicitation for rational agent. In Dirk Reichardt, Paul Levi, and John-Jules C. Meyer, editors, *Proceedings of the 1st workshop on Emotion and Computing - Current Research and Future Impact*, Bremen, Germany, June 2006.

[182] Magalie Ochs, Radoslaw Niewiadomski, Catherine Pelachaud, and David Sadek. Intelligent expressions of emotions. In *First International Conference, ACII 2005, Beijing, China, October 22-24, 2005*, volume 3784/2005, pages 707–714. Springer Berlin / Heidelberg, 2005.

[183] Eugénio Oliveira and Luís Sarmento. Emotional advantage for adaptability and autonomy. In *Proceedings of the Second International Joint Conference on Autonomous Agents & Multiagent Systems*, pages 305–312, Melbourne, Australia, 2003.

[184] O. OŃeill. *Autonomy and Trust in Bioethics.* Cambridge University Press, Cambridge, 2002.

[185] OnSale. Onsale. http://www.onsale.com, 2010.

[186] Andrew Ortony, Geraldl L. Clore, and Allan Collins. *The Congnitive Structure of Emotions*. The Cambridge University Press, 1988.

[187] T. Otsuka and J. Ohya. Recognizing multiple persons' facial expressions using HMM based on automatic extraction of significant frames from image sequences. In *Proceedings of the International Conference on Image Processing*, pages 546–549, 1997.

[188] Oxygen. Project oxygen. http://www.oxygen.lcs.mit.edu/, 2004.

[189] J. Panksepp. Basic emotions ramify widely in the brain, yielding many concepts that cannot be distinguished unambiguously. In P. Ekman and R.J. Davidson, editors, *The nature of emotion*, pages 86–88. Oxford University Press, Oxford, England, 1994.

[190] B. Parkinson, P. Totterdell, R.B. Briner, and S. Reynolds. *Changing moods: The psychology of mood and mood regulation*. Addison Wesley Longman, Harlow, England, 1996.

[191] H. Van Dyke Parunak, Robert Bisson, Robert Matthews Sven Brueckner, and John Sauter. A model of emotions for situated agents. *AAMAS'06*, pages 993–995, May 2006.

[192] Vladimir I. Pavlovic, Rajeev Sharma, and Thomas S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):677–695, 1997.

[193] Catherine Pelachaud. Studies on gesture expressivity for a virtual agent. *Speech Commun.*, 51(7):630–639, 2009.

[194] V. Petrushin. How well can people and computers recognize emotions in speech? In *Proceedings of the AAAI Fall Symposium*, pages 141–145, 1998.

[195] R. W. Picard. *Affective computing*. MIT Press, 1997.

[196] Ingmar Porn. *Action Theory and Social Science*. Synthese Library 120. D. Reidel Publishing Company, Holland, 1977.

[197] Ingmar Porn. On the nature of a social order. In J. E. Fenstad et al., editor, *Logic, Methodology and Philosophy of Science*, pages 553–567. Elsevier Science Publishers, 1989.

[198] S. Poslad, M. Calisti, and P. Charlton. Specifying standard security mechanisms in multi-agent systems. In *Proceedings of the Workshop on Deception,*

Fraud and Trust in Agent Societies, AAMAS 2002, pages 122–127, Bologna, Italy, 2002.

[199] M. Power and T. Dalgleish. *Cognition and Emotion: From order to disorder.* Psychology Press, Taylor & Francis, Hove and New York, 1997.

[200] G. Priest and K. Tanaka. *Paraconsistent Logic.* Stanford Encyclopedia of Philosophy, 1996.

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

[202] S. D. Ramchurn, C. Sierra, L. Godo, and N. R. Jennings. Devising a trust model for multi-agent interactions using confidence and reputation. *International Journal of Applied Artificial Intelligence*, 18(9–10):833–852, 2004.

[203] Sarvapali D. Ramchurn, Dong Huynh, and Nicholas R. Jennings. Trust in multi-agent systems. *The Knowledge Engineering Review*, 19(1):1–25, 2004.

[204] P. V. Rangan. An axiomatic basis of trust in distributed systems. In *Proceedings of the Symp. Security and Privacy*, Washington, DC, 1988. IEEE Computer Society Press.

[205] Rainer Reisenzein. Emotions as metarepresentational states of mind: Naturalizing the belief-desire theory of emotion. *Cognitive Systems Research*, 10(1):6–20, 2009.

[206] Sebastian Ries, Jussi Kangasharju, and Max Muhlhauser. A classification of trust systems. In R. Meersman, Z. Tari, and P. Herrero et al., editors, *OTM Workshops 2006, LNCS 4277*, pages 894–903. Springer-Verlag Berlin Heidelberg, 2006.

[207] M. Roman, C. Hess, R. Cerquiera, A. Ranganathan, R. H. Campbell, and K. Nahrstedt. A middleware infrastructure for active spaces. *IEEE Pervasive Computing*, 1:74–83, 2002.

[208] I. Roseman and C. Smith. Appraisal theories: Overview, assumptions, varieties, controversies. In A. Schorr K. R. Scherer and T. Johnstone, editors, *Appraisal processes in emotion: Theory, Methods, Research*, pages 3–19. Oxford University Press, 2001.

[209] I.J. Roseman. Cognitive determinants of emotion: a structural theory. In P. Shaver, editor, *Review of Personality and Social Psychology: Emotions, Relationships, and Health*, chapter 1, pages 11–36. Sage, Beverly Hills, CA, 1984.

[210] I.J. Roseman. Appraisal determinants of discrete emotions. *Cognition and Emotionn*, 5(3):161–200, 1991.

[211] M. Rosenblum, Y. Yacoob, and L. Davis. Human expression recognition from motion using a radial basis function network architecture. *IEEE Trans. on Neural Network*, 7(5):1121–1138, 1996.

[212] J. B. Rotter. A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35(4):651–665, 1967.

[213] Daniel Rousseau and Barbara Hayes-Roth. A social-psychological model for synthetic actors. In Katia P. Sycara and Michael Wooldridge, editors, *2nd International Conference on Autonomous Agents (Agents'98)*, pages 165–172. ACM Press, New York, 1998.

[214] Sini Ruohomaa and Lea Kutvonen. Trust management survey. In *Procceedings of iTrust 2005, number 3477 in LNCS*, pages 77–92. Springer-Verlag, 2005.

[215] Jordi Sabater and Carles Sierra. REGRET: A reputation model for gregarious societies. In *Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies*, pages 61–69, Montreal, Canada, 2001.

[216] Jordi Sabater and Carles Sierra. Reputation and social network analysis in multi-agent systems. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-02)*, pages 475–482, Bologna, Italy, July 15–19 2002.

[217] Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artif. Intell. Rev.*, 24(1):33–60, 2005.

[218] Kathryn E. Sanders. A logic for emotions: A basis for reasoning about commonsense psychological knowledge. Technical report, Providence, RI, USA, 1989.

[219] Stanley Schacter and Jerome Singer. Cognitive, social and physiological determinants of emotional states. *Psychological Review*, 69:379–399, 1962.

[220] K. Scherer. Adding the affective dimension: A new look in speech analysis and synthesis. In *Proceedings of the International Conference on Spoken Language Processing*, pages 1808–1811, 1996.

[221] K. R. Scherer. Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99:143–165, 1986.

[222] K. R. Scherer. *Appraisal Processes in Emotion : Theory, Methods, Research*, chapter Appraisal Considered as a Process of Multilevel Sequential Checking, pages 92–120. Oxford University Press, New York, 2001.

[223] M. Schillo, P. Funk, and M. Rovatsos. Using trust for detecting deceitful agents in artificial societites. *Applied Artificial Intelligence (Special Issue on Trust, Deception and Fraud in Agent Societies)*, 2000.

[224] Ph. Schnoebelen. The complexity of temporal logic model checking. In Philippe Balbiani, Nobu-Yuki Suzuki, Frank Wolter, and Michael Zakharyaschev, editors, *Advances in Modal Logic 4, papers from the fourth conference on "Advances in Modal logic," held in Toulouse (France) in October 2002*, pages 393–436. King's College Publications, 2002.

[225] M. Schröder and S. Breuer. XML representation languages as a way of interconnecting TTS modules. In *Proceedings of the 8th International Conference on Spoken Language Processing (INTERSPEECH 2004 - ICSLP)*, Jeju Island, Korea, October 2004.

[226] Norbert Schwarz and Gerald L. Clore. Mood, misattribution, and judgments of weil-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45(3):513–523, 1983.

[227] John R. Searle. *Intentionality (Cambridge Paperback Library).* Cambridge University Press, December 2003.

[228] S. Sen and N. Sajja. Robustness of reputation-based trust: Booblean case. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-02)*, pages 288–293, Bologna, Italy, 2002.

[229] Murat Sensoy, Jie Zhang, Pinar Yolum, and Robin Cohen. Poyraz: Context-aware service selection under deception. *Computational Intelligence*, 25(4):335–366, 2009.

[230] S. P. Shapiro. The social control of impersonal trust. *American Journal of Sociology*, 93(3):623–658, 1987.

[231] Yoav Shoham. *Reasoning about Change: Time and Causation from Standpoint of Artificial Intelligence.* MIT Press, Cambridge, MA, 1988.

[232] Grigori F. Shvarts. Gentzen style systems for K45 and K45D. In *Proceedings of the Symposium on Logical Foundations of Computer Science*, pages 245–256, London, UK, 1989. Springer-Verlag.

[233] J. Sichman, R. Conte, C. Castelfranchi, and Y. Demazeau. A social reasoning mechanism based on dependence networks. In *Proceedings of the 11th ECAI*, 1994.

[234] Robert Solomon. Emotions and choice. In Amélie Rorty, editor, *Explaining Emotions*, pages 251–281. University of California Press, Los Angeles, 1980.

[235] Stanford. *Emotion.* Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/emotion/, 2010.

[236] Stanford. *Trust.* Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/trust/, 2010.

[237] Lucas Stephane. Cognitive and emotional human models within a multi-agent framework. In *Proceedings of the International Conference of Human - Computer Interface*, Beijing, China, 2007.

[238] Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. 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*, pages 142–147. AAAI Press, 2007.

[239] Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer. Towards a quantitative model of emotions for intelligent agents. In Dirk Reichardt and Paul Levi, editors, *Proceedings of the 2nd Workshop on Emotion and Computing - Current Research and Future Impact*, Osnabrück, Germany, 2007.

[240] Bas R. Steunebrink, Mehdi Dastani, and John-Jules Ch. Meyer. A formal model of emotion-based action tendency for intelligent agents. In Luís Seabra Lopes, Nuno Lau, Pedro Mariano, and Luís Mateus Rocha, editors, *Progress in Artificial Intelligence, 14th Portuguese Conference on Artificial Intelligence, EPIA 2009, Aveiro, Portugal, October 12-15, 2009*, volume 5816 of *Lecture Notes in Computer Science*, pages 174–186. Springer, 2009.

[241] Sun. X.509 certificates and certificate revocation lists (crls). Technical report, http://java.sun.com/products/jdk/1.2/docs/guide/security/cert3.html.

[242] Jianhua Tao. Emotion control of chinese speech synthesis in natural environment. *Eurospeech2003*, 2003.

[243] Gabriele Taylor. Justifying the emotions. *Mind*, 84:390–402, 1975.

[244] W. T. Teacy, Jigar Patel, Nicholas R. Jennings, and Michael Luck. Travos: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198, 2006.

[245] Bill Tomlinson. The logic of emotion. http://homepage.mac.com/billtomlinson/LOE3.html, 2001.

[246] Dinh Que Tran and Manh Hung Nguyen. Modeling trust in open distributed multiagent systems. In *Proceeding of the International Conference in Mathematics and Applications (ICMA-MU 2009), Bangkok, Thailand, December 2009*, 2009.

[247] Johan van Benthem. Correspondence theory. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 3, pages 325–408. Kluwer Academic Publishers, 2 edition, 2001.

[248] Johan van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 14(2), 2004.

[249] W. van der Hoek, B. van Linder, and J.-J. Ch. Meyer. An integrated modal approach to rational agents. In M. Wooldridge and A. Rao, editors, *Foundations of Rational Agency*, pages 133–168. Applied Logic Series 14, Kluwer, Dordrecht, 1998.

[250] Yde Venema. Temporal logic. In Lou Goble, editor, *The Blackwell Guide to Philosophical Logic*. Blackwell, 2001.

[251] Patricia Victor, Chris Cornelis, Martine De Cock, and Paulo Pinheiro da Silva. Gradual trust and distrust in recommender systems. *Fuzzy Sets and Systems*, 160(10):1367–1382, 2009. Special Issue: Fuzzy Sets in Interdisciplinary Perception and Intelligence.

[252] Georg Henrik von Wright. *An Essay in Modal Logic*. North-Holland publishing Co., Amsterdam, 1951.

[253] D. N. Walton and E. C. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New-York Press, NY, 1995.

[254] Yonghong Wang and Munindar P. Singh. Formal trust model for multiagent systems. In *IJCAI'07: Proceedings of the 20th international joint conference on Artifical intelligence*, pages 1551–1556, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.

[255] B. Weiner. An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4):548–573, 1985.

[256] A. Wells and G. Matthews. *Attention and Emotion: A clinical perspective.* Psychology Press, Taylor & Francis, Hove UK, 1994.

[257] William Whisner. A new theory of emotion: Moral reasoning and emotion. *Philosophia*, 31(1-2):3–30, 2003.

[258] C. Williams and K. Stevens. Emotions and speech: Some acoustical correlates. *Journal of the Acoustic Society of America*, 52(4):1238–1250, 1972.

[259] O. E. Williamson. Calculativeness, trust, and economic organization. *Journal of Law and Economics*, 34:453–502, 1993.

[260] Y. Yacoob and L. Davis. Recognizing human facial expressions from long image sequences using optical flow. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(6):636–642, 1996.

[261] Chengwei Yao and Gencai Chen. A emotion development agent model based on OCC model and operant conditioning. In *Proceedings of 2001 International Conferences on Info-tech and Info*, volume 3, pages 246–250, China, 2001.

[262] B. Yu and M. P. Singh. Towards a probabilistic model of distributed reputation management. In *Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies*, pages 125–137, Montreal, Canada, 2001.

[263] B. Yu and M. P. Singh. Distributed reputation management for electronic commerce. *Computational Intelligence*, 18(4):535–549, 2002.

[264] B. Yu and M. P. Singh. An evidential model of distributed reputation management. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-02)*, pages 294–301, Bologna, Italy, 2002.

[265] Feng Yu, Eric Chang, Yingqing Xu, and Heung-Yeung Shum. Emotion detection from speech to enrich multimedia content. In *PCM '01: Proceedings of the Second IEEE Pacific Rim Conference on Multimedia*, pages 550–557, London, UK, 2001. Springer-Verlag.

[266] G. Zacharia. Collaborative reputation mechanisms for online communities. Master's thesis, Massachusetts Institute of Technology, 1999.

[267] Jiehan Zhou, Changrong Yu, Jukka Riekki, and Elise Kärkkäinen. AmE framework: a model for emotionaware ambient intelligence. In *Proceedings of the second International Conference on Affective Computing and Intelligent Interaction (ACII2007): Doctoral Consortium.*, Lisbon, Portugal, 2007.