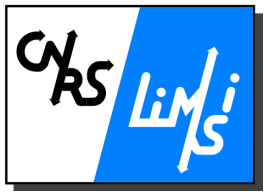


Master of Research - Intelligent and Communicating Systems

Master Thesis Report

# A Logical Model of Affective and Interaction-Oriented Theory of Mind


By

Marwen BELKAID

ETIS - ENSEA / Université de Cergy-Pontoise / CNRS UMR 8051
6 avenue du Ponceau, 95014 Cergy-Pontoise Cedex, France


Presented on the 19th of September of 2013

In the presence of the following Jury Members:

| | | | |
|---|---|---|---|
| Pr. | Ph. GAUSSIER | Université de Cergy-Pontoise | Chair of the Jury |
| A.Pr. | A. PITTI | Université de Cergy- Pontoise | Reader |
| Pr. | N. SABOURET | Université Paris-Sud | Supervisor |
| Pr. | J.C. MARTIN | Université Paris-Sud | Supervisor |

# *Aknowledgments*

*I am very grateful to the LIMSI-CNRS Laboratory for accepting me within their researchers' team and for providing all necessary conditions for my internship.*

*I would like to express my sincere appreciation to Pr. Nicolas SABOURET and Pr. Jean-Claude MARTIN for giving me the opportunity to work on such an interesting project. I am very thankful for the attention, time and support they granted me during the whole training period.*

*I also wish to extend my thanks to the CPU, AMI and TARDIS teams and all the people I met via this project. My gratitude is particularly reserved to Ph.D. Hazaël JONES for his guidance concerning the integration in the TARDIS project as well as to A.Pr. Céline CLAVEL, Caroline FAUR, Tom GIRAUD and Leonor PHILIP for their advices and help regarding the evaluation study.*

*I finally take this opportunity to thank all the academic and administrative staff involved in the Intelligent and Communicating Systems Research Master's degree.*

A tous ceux qui mériteraient que je leur dédie ce travail.
A mon étoile.

# Contents

# List of Figures

# List of Acronyms

| | |
|---|---|
| *AI* | Artificial Intelligence |
| *ANOVA* | ANalysis Of VAriance |
| *API* | Application Programming Interface |
| *BDI* | Beliefs Desires Intentions [model] |
| *DLL* | Dynamic Link Library |
| *IVA* | Intelligent Virtual Agent |
| *OCC* | Ortony, Clore, and Collins [emotions theory] |
| *ST* | Simulation-Theory |
| *TT* | Theory-Theory |
| *Tom* | Theory of Mind |
| *XML* | eXtensible Markup Language |

# Introduction

The concept of *Theory of Mind* (ToM) was first introduced in 1978 by Premack and Woodruff in their paper "Does the chimpanzee have a theory of mind?" as the ability to attribute mental states to oneself and others. Also often referred to by the term *mindreading*, it defines how human – and eventually non-human – beings interpret, explain and predict their own and others' behavior in terms of goals and intentions. In other word, how they theorize about their own and others' mind.

This ability is commonly, and often unconsciously, used by human beings and plays a fundamental role in their social interaction. Therefore, it has been widely investigated for decades and in various research disciplines. Psychologists mainly focus on the issue of testing its existence and its development during childhood [Wellman 90] [Wimmer 83]. The modular approach, for example, is based on the idea of innate modules in human brain that are involved in mindreading [Leslie 94] [Baron-Cohen 97]. Besides, philosophers are interested in the kind of processes on which the theory of mind in adults relies. This has led to a debate, still in progress, between *theorists* – arguing in favor of a folk-psychology reasoning – and *simulationists* – defending a projection or a mirroring process [Botterill 99] [Goldman 06]. Moreover, neuroscientists investigate the brain regions that are involved when it comes to reason about one's own and others' minds [Vogeley 01].

*Affective Computing* is an interdisciplinary field that examines the development of computer systems that can recognize, interpret and simulate human emotions. It presents challenges both in the creation of more powerful and "user-friendly" technologies and in the construction of computational models that allow for testing theories about human emotions and behavior. One of the branches of this field aims at the implementation of Intelligent Virtual Agents (IVA) that would be able to interact, not only with each other in multiagents systems but also with human users. However, according to [Castelfranchi 97], in order to understand and collaborate with the latter, agents need be social. Moreover, "sociality" must not be reduced to communication but rather it would have to encompass it, along with other attitudes such as cooperation, competition, delegation, manipulation, that often rely on mindreading processes.

In this work, we investigate the contribution of an *affective theory of mind* in Human/Agent *interaction*. Thus, we propose a non-domain-specific theoretical model that

gives IVAs the ability to reason about the user's mental and emotional states. We argue that such a module represents a step toward the enrichment of the agent's social behavior and an enhanced realism of interactions.

**Chapter 1** of this report presents a state of the art related to the key notions of our project, e.g. theory of mind, emotion, interaction, etc. In **Chapter 2**, we introduce the logical framework which we propose for interaction-oriented affective ToM. Then, **Chapter 3** describes the global architecture intended to encompass a reasoning engine based on it and gives details about our implementation of this theoretical model. Finally, in **Chapter 4**, we tackle the evaluation of this model through a set of experiments.

# Project context

This project is conducted in the context of an intership Human/Machine Communication department of *Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur* (LIMSI) laboratory, within the Architectures and Models for Interaction (AMI) and the Cognition, Perception and Usages (CPU) teams. AMI studies the interactional phenomena that occur between humans and computer systems and examines the development of new interaction paradigms. On the other hand, CPU explores the cognitive, perceptual and emotional processes in human and virtual agents and addresses research topics such as perceptual systems and models and emotion in virtual agents. See LIMSI web page[1] for more details.

This project is also highly connected to the TARDIS project. The TARDIS consortium is composed of research teams from european academic organisms as well as private commercial and non-commercial parteners. The project aims at the development of an open-source platform for online and offline social training for young people at risk of social exclusion. It mainly addresses the case of job interviews and intends to facilitate youngsters' access to employment. Therefore, this is also the main application context we will focus on. In TARDIS general architecture, our work is part of the *Affective Module* that reasons about the user's mental state and the course of the interaction. See TARDIS web page[2] for more details about the project and Section 3.4 for information about the connection between with our module.

The context of the project we presented here also defines our topics of research which are Human/Machine interaction enhancement and cognitive and emotional processes modeling in virtual agents.

[1] http://www.limsi.fr/Scientifique/index.en.html
[2] http://tardis.lip6.fr/

# Chapter 1

# Bibliographical study

The purpose of our work is to investigate the contribution of an affective theory of mind in Human/Agent interaction. Thus, **Section 1** of this chapter introduces some ToM theories and models regarding its development in children and its functioning in adults. This will allow for better understand this concept. In **Section 2**, we consider the Beliefs-Desires-Intentions (BDI) theory and enumerate some benefits from using such a model to represent agents' mental states. Next, **Section 3** focuses on the emotions' perspective and describe the models that we rely on for affects appraisal and triggering processes. Finally, **Section 4** tackles Human/Agent communication and interaction. Along the sections, we also have an overview of the related work and discuss our approach compared to other projects.

## 1.1 Theory of Mind modeling

### 1.1.1 Development of the Theory of Mind in infants

#### From the psychologists perspective

Leslie considers the development of pretense ability in 2-year-old children as the first step in understanding cognition and, consequently, as an "early manifestation" of the theory of mind. "Pretending oneself is thus a special case of the ability to understand pretense in others" [Leslie 87]. Pretend representations, which he defines as *meta-representations*, i.e. representations of representations – as opposed to primary ones which are intended to be accurate representations of the real world – are thus the key connexion between pretense and mindreading [Leslie 94]. Besides, Leslie proposed a model of ToM in which the representation of causal events is central. According to him, there are three modules that deal with the different classes of these events: 1) ToBY (Theory of Body) for events

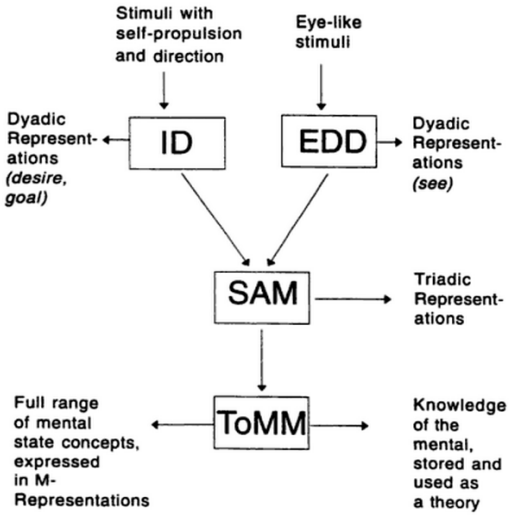



Figure 1.1: Baron-Cohen's model of Theory of Mind [Baron-Cohen 97]

that are described by the rules of mechanics (*mechanical agency*), e.g. Y moved because X pushed it, 2) ToMMs1 (Theory of Mind Mechanism (system 1)) for events that are described in terms of intents, goals and actions (*actional agency*), e.g. X goes to the kitchen in order to eat, and 3) ToMMs2 (Theory of Mind Mechanism (system 2)) for events that are described in terms of attitudes and beliefs (*attitudinal agency*), e.g. X opens the fridge because he thinks there is food inside. The latter is the one that employs meta-representations for reasoning about one's own and others' mind [Leslie 94].

Baron-Cohen tackled mindreading processes in his study of autism and social capabilities of non-human primates and other vertebrates [Baron-Cohen 97]. In his model of ToM, he defines four modules that are used in the mindreading system: 1) the Intentionality Detectors (ID) interprets self-propelled motions in terms of goals and desires and distinguishes animate stimuli from objects (A wants B), 2) the Eye Direction Detector (EDD) determines direction of gaze based on the detection of eye-like visual stimuli (C sees D), 3) the Shared Attention Mechanism(SAM) uses the dyadic information from ID and EDD in order to produce nested (triadic) representations interpreting eye direction in terms of goals (C sees (A wants B)), 4) the Theory of Mind Mechanism (ToMM) produces metarepresentations to express mental states based on one's experience. See Figure 1.1.

In Leslie's model, ToBy starts developing in the first months, followed by ToMMs1 around the age of 6 months and then ToMMs2 the 18th and 48th months [Leslie 94]. In Baron-Cohen's model, ID and EDD emerge in the first 9 months while SAM develops between 9 and 18 months and ToMM between 18 and 48 months [Baron-Cohen 97]. These *modular theories* both match the developmental progression that is observed in normal infants and are consistent with others studies. One of the most important step in the development of mindreading ability is the attribution of *false belief*, i.e. understanding that

someone might hold a representation that is different from the real world or from ours. Wimmer and Perner conducted the first systematic investigation on this capacity and it turned out that children failed the false belief task until the age of 4 years [Wimmer 83]. Furthermore, [Wellman 90] shows that 2-year-old children only interpret event in terms of action, unlike older children and adults that are able to understand them in terms of desires and beliefs.

Although we will not tackle the learning phase of mindreading in our work, information on the origins and development of ToM help us better understand its basis and might, in addition, be useful for future work .

#### Implementations

In [Scassellati 02], an application of Leslie's and Baron-Cohen's models of ToM on a humanoid robot is presented. This initial implementation focuses on the basic skills such as faces and eyes detection, discrimination of animate and inanimate and gaze following. A module of "partial" ToM is also presented in [Peters 05]. Here, the detection of gaze and intentionality, based on Baron-Cohen's model, is used for conversation initiation in virtual agents: depending on internal goals and on the attention that other agents pay to them, agents decide whether they engage in interaction.

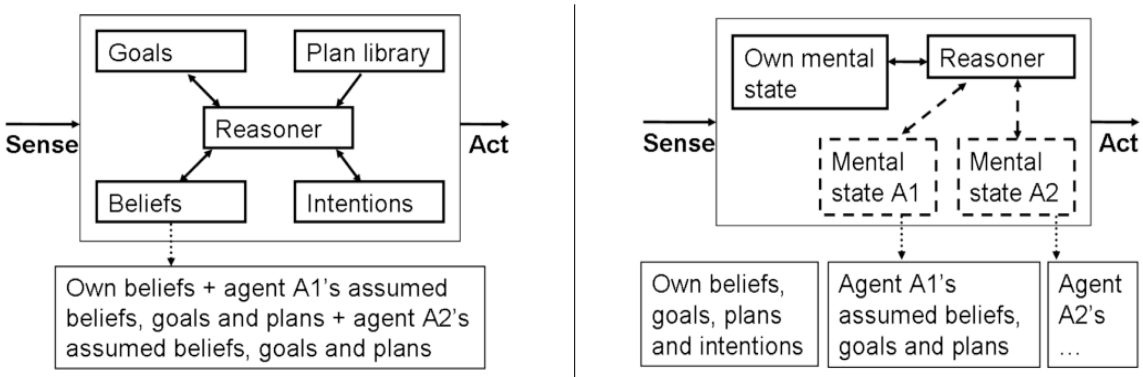
However, the aim of our project is to focus on the cognitive processing of ToM, i.e. respectively ToMMs2 and ToMM in Leslie's and Baron-Cohen's models, rather than to handle the "lower-level" modules of ToM. We do not need these features for conversation initiation either, since, in our case, the interactions are based on scenarios.

### 1.1.2 Human adults Theory of Mind processing

#### Two philosophical theories

While psychologists mainly focused on the development of mindreading in young children, a philosophical debate has been run about how ToM was processed by adults. This debate opposes two theories: the *theory-theory* and the *simulation-theory* [Botterill 99] [Goldman 06] [Vogeley 01] [Harbers 11].

The theory-theory (TT) is based on the so-called *folk psychology* or commonsense that refers to how people think they think, i.e. their theory of the functioning of human mind. Beliefs, desires and intentions are thus imputed to others by intuition and then a set of principles, about how theses mental states interact with each other, is used to understand their behaviors. This implies that rules about others' behavior would be held in the agent's knowledge base in order to be used by the ToM module.




(a) TT (b) ST

Figure 1.2: Harbers' architectures for TT and ST models of Theory of Mind [Harbers 11]

On the other side, simulation-theory (ST) states that ToM is the capacity to mimic other people mental states and to project one's own attitutude on them. Human then take someone else's perspective and use their own reasoning capacities to interpret the observed events. The existence evidence of a "mirror neural system" that is activated both when executing an action and when observing someone else is executing it, in macaques and in human, gave this theory a significant support. This theory suggests that the agent would use its own reasoner or inference engine in order to simulate others' reactions.

### Implementations

In [Harbers 11], two ToM models, based respectively on Simluation-Theory (ST) and Theory-Theory (TT), are implemented for the purpose of virtual training. Simulation studies demonstrated a higher performance in agents having ToM compared to those who do not and regarding to the expected behavior. But, there were no difference between ST and TT implementations in these results. Nevertheless, ST implementation turned out to be better from the developper's point of view because of code reusability and flexibility in mental states modification.

### Towards a hybrid theory

Since these theories were introduced, various research demonstrated that a pure TT or ST was not realistic. For instance, according to [Vogeley 01], ToM and self-perspective (SELF) cognitive processes rely on both common (anterior cingulate cortex and right prefrontal cortex) and differential (left temporopolar cortex for ToM and right temporoparietal junction for SELF) neural mechanisms. Those results consequently reject simulation-theory and theory-theory concepts and suggest a mixture of both theories. Indeed, more moder-

ate theories on ToM functionning appeared. On the one hand, *Weakened TT* accepts that mindreading also involves simulation although the central role is still played by folk psychology [Botterill 99] [Nichols 03]. On the other hand, the *hybrid simulationist approach* clames that the mirroring functions remain central (e.g. neural emotion centers activated by the recognition of facial expressions) even though it concedes an important role to theory in predicting and explaning someone else's action [Goldman 06]. The main difference between the two theories is probably whether children learn to attribute mental states to others at the same time or after they learn to attribute them to themselves.

In any case, in our work, we will adopt a hybrid approach where part of the Theory of Mind reasoning will rely on folk-psychology mental states representation and commonsense rules held in the knowledge base. On the other hand, a projection process will be used whenever an assumption can be made that others have similar inference engines, which reduces the development cost as pointed out by [Harbers 11] or in unknown situations where the agent's knowledge and rules are not sufficient.

## 1.2 Beliefs, Desires and Intentions as core mental states

### Theoretical support

The BDI model is a well-known and very common model used in human behavior representation and intelligent agents development [Rao 91] [Bosse 11] [Harbers 11]. It bases the interpretation and understanding of human practical reasoning on three core attitudes: Beliefs, Desires and Intentions (BDI). It implements Bratman's theory, which is significantly based on folk psychology. The particularity of his theory is that the intention is treated as a crucial element of practical reasoning. It does not only characterizes the action but also the mind. It is the partial action plan that someone is committed to achieve to fulfill his/her goal [Bratman 99].

This model is also consistent with other theories in psychology in which people mostly use reason explanation – based on beliefs and desires – for intentional behavior, where intention mediates between reason and actions [Malle 99] [Wellman 90] [Wimmer 83]. Therefore, we do believe BDI theory is appropriate for human cognition modeling.

### Formal BDI model

Rao and Georgeff proposed a formalization of BDI theory aimed to be used in intelligent agents modeling [Rao 91]. As they say, their formalism is similar to Computation Tree

Logic CTL*. A possible world is symbolized by a time tree with a single past and branches to illustrate the choices and the events. The temporal operators are *next*, *eventually*, *always* and *until*. Then, they combine this temporal logic with modal logic using two modalities: *optional* and *inevitable* [Rao 91].

Beliefs correspond by the possible worlds, i.e. distinct time trees with probabilities of occurence, and can be seen as the informative component of the system. Desires represent its motivational state and can be inconsistent with one another. However, goals are chosen among those desires and have to be consistent with one another and believed to be achievable. Finally, intentions represent the deliberative component, that is to say the paths selected by the deliberation function as the best regarding to the goals. As it represents a partial action plan, this additional component is used to obtain a balance between reactive (no plan) and goal-directed (one plan) behavior. The agent thus commits to its plans yet periodically reconsiders them given new states of affairs [Rao 91][Rao 95].

### BDI Vs non-BDI Theory of Mind models

PsychSim is a simulation tool modeling interaction between agents that have a ToM. Mental states are represented using the COM-MTDP framework, instead of the BDI model, in order to address two shortcomings of the latter in the context of decision problems in multi-agent systems: the lack of characterization of computational complexity of teamwork decisions and the absence of techniques for quantitative evaluation of optimality degree [Pynadath 02] The agents then have a fully specified decision-theoretic model of their environment, including beliefs about the world and recursive models of other agents. Behaviors are only represented in terms of beliefs and desires [Pynadath 05]. Nevertheless, our work deals with Human/Agent interaction rather than multi-agent interaction. Hence, we are not concerned with the BDI shortcomings mentioned above.

[Bosse 11] presents a BDI-based model for mindreading in which folk psychology intentional stance is taken as a point of departure. ToM operates in two levels. The first one allows for social anticipation, i.e. predicting others' behavior in advance. The second one allows for social manipulation, i.e. trying to affect the occurence of certain mental states in advance. The model is tested in three application areas: social manipulation, animal cognition and virtual storytelling. [Harbers 11] model is also based on BDI, but unlike the two previous mentioned projects that use a TT approach, it implements both TT and ST. Though, because radical simulationists claim that attitudes are not necessarily represented in the way folk-psychology says, in the ST model attributed mental states are not expressed in BDI but translated to it before they are processed by the reasoner.

In our case, for several reasons, we believe that the BDI model is appropriate for human cognition, and especially mindreading, modeling: a strong theoretical support, a clear mental states representation, a large formalization and implementation literature, etc.

Hence, the hybrid approach we adopt will rely on its representation of human mind and deliberation process, although we do not intend to take part in the TT Vs ST debate. Furthermore, whereas the examples mentioned above are very related to our work, none of them considers to affective dimension of human interactions in their models.

## 1.3 An affective aspect in mindreading

**Emotion theories in psychology**

Modern psychology and neuroscience attribute a significant role to emotions in deliberation. And given the large range of theories proposed by psychologists, modeling agents in affective computing requires a study of the conceptual and theoretical background to rely on. [Scherer 10] presents a survey of emotion theories and some criteria to take into account when designing *emotionally competent* agents. In this work, our purpose is to model a ToM that allows for reasoning about others' belief and goals as well as their affective states. Therefore, we are interested in a cognitive approach for emotions.

In appraisal theories, emotions are generated by an evaluation, namely an *appraisal*, of events or, more generally, states of affairs that determines the reaction within different *coping* strategies [Scherer 10]. According to Arnold, the appraisal process is distributed over several components: physiological reaction (hormonal mechanisms), motor reaction (facial expression, gesture, etc.), motivation for actions (running, jumping, etc.) and subjective feeling (determining the name we give to the emotion for instance) [Arnold 60]. In Scherer's Component Process Model (CPM), the evaluation of internal and external stimuli elicits changes in the states of all or most of the five components he defines as *organismic subsystems* in the form of sequential checks [Scherer 01]. Furthermore, Lazarus considers two levels of cognitive appraisal for events: the *primary* one evaluate them with regard to one's goals and the *secondary* one with regard to one's adaptation to their consequences. The way humans cope with these events depends on how this evaluation alters their mental states, i.e. their beliefs, desires, intentions, etc., as well as on their personality. Humans also have the ability to influence their reaction according to various coping strategies, e.g. problem-focused coping in which actions are engaged and emotion-focused coping in which one attempts to influence the emotional response [Lazarus 91].

Ortony, Clore, and Collins proposed a semi-formal description of emotions and their cognitive structure [Ortony 90]. The OCC theory thus distingues twenty-two types of emotions divided in three main branches as reaction to one of the following stimuli kind: consequences of events, actions of agents, and aspects of objects. These stimuli are assessed under a central criterion – the central appraisal variable – that evaluate those stimuli in terms of desirability of an event, approbation of an action and attraction of an object. Secondary appraisal variables, such as the likelihood, the unexpectedness or the praise-

worthiness of a state of affairs, influence the intensity of the triggered emotion. With their theory, the authors wanted to provide an easily understandable and computationally tractable model of emotion that could be used in artificial intelligence [Ortony 90].

Because of its simplicity, its implementability and its compatibility with BDI-based models, OCC appears to be an appropriate theory to rely on in our model.

### Formalizations and implementations of the OCC model

A formalization of OCC theory is presented in The authors use BDI logic, relying on the strong connection between cognition and emotions in the sense that they are both based on mental states such as beliefs, goals and intentions. Indeed, what is called emotions here are the *Intentional* affective states, i.e. affects that are about or directed to something. This formalization of twenty of the twenty-two emotions of OCC theory provides a good starting point for our model. However, the logical framework uses a lot of different operators, some of which we will not need in this project. Moreover, this formalism ignores the quantitative aspect of emotions, i.e. their intensity, as well as beliefs and desires levels for instance.

[Dastani 12] provides a logical framework in which the authors consider emotions from appraisal to coping. The effect of the appraised situation on those mental states, and hence on the behavior of the agent, is determined by the emotions intensity and the selected coping strategies. The value of an emotion depends on the corresponding levels of believability and desirability. These are symbolized by the *graded beliefs* and *goals*. While it gives an interesting response to one of the issues we pointed out above, this model still does not provide all the modalities we need to model social relations and interactions as we aim to. On the contrary, similarly to [Adam 09], it models some aspects we are not interested in, like preconditions for action selection or general probability or *exceptionality* of events.

FatiMA is an affective agent architecture in which Theory of Mind is considered from the emotions perspective [Aylett 08]. Based on Simulation-Theory and OCC cognitive taxonomy of short-term affects, the existing appraisal mechanism is used to predict the emotional response to the set of actions the agent could possibly take. This *double appraisal* mechanism allows for making a virtual drama actor assess the emotional effect of its behavior on its audience in order to generate more interesting emergent narratives. Although FatiMA architecture is not BDI-based, it defines internal states such as knowledge, goals and intentions. Nevertheless, to assess others' potential reaction, FatiMA agents only uses their actual emotional state. Even though it is admitted that the character cannot assume that the others are the same as him, sharing its beliefs and goals, this model do not take into account their own mental states. Besides, the agent's objective is to produce the most dramatical effect and to induce the greatest emotional impact, which is less relevant in other kinds of applications. The way emotion-focused goals can be handled in more general contexts is not considered. Contrariwise, in our work, we aim to investigate the

effect of a theory of mind model in any kind of Human/Agent interactions.

## 1.4 Communication and social interaction in intelligent agents

### 1.4.1 Speach acts theories

Austin's speech acts theory distinguishes three levels in the act of speaking: 1) locutionnary acts, referring to the produced sound, the lexical and grammatical conventions and the surface meaning of an utterance, 2) illocutionnary acts, referring to its intented social meaning (e.g. asking a question), and 3) perlocutionnary acts, referring to its actual effect (e.g. eliciting an answer) [Austin 62].

In Searle's theory, which is highly inspired by Austin's, what characterizes an illocution is the *meaning* of the utterance according to the set of rules and conventions of the used language [Searle 69]. He distinguishes four types of rules: 1) *propositional content* rules, 2) *preparatory* rules, 3) *sincerity* rules and 4) *essential* rules. For instance, when $S$ promises $H$ that $p$, 1) the utterance predicates a future act $A$ of $S$, 2) $S$ believes that $H$ would prefer $S$'s doing A than his not doing it and it is not obvious to both that $S$ would do $A$ in the normal course of actions, 3) $S$ intends to take responsibility for intending to do $A$, and 4) the utterance counts as the undertaking of an obligation to do A. Thus, the illocutionary *act* results from the intention to produce an illocutionary *effect*: that the hearer recognizes that the states of affairs specified by some of the rules obtain [Searle 69].

Moreover, Searle's taxonomy divides illocutionnary acts in five basic classes: 1) assertives, i.e. stating facts and expressing *Beliefs*, 2) directives, i.e. describing orders or requests and expressing *Desires*, 3) commissives, i.e. representing commitment and expressing *Intentions*, 4) expressives, i.e. describing and expressing *Emotions*, and 5) declarations, i.e. modifying reality [Searle 76] [Searle 69].

#### Formalizations and implementations

Because of their role in representing psychological attitudes for intelligent agents, many formalizations of the illocutionnary acts are available in the literature. [Herzig 02] presents an interesting framework based on Beliefs and Intentions where assertives are the basis of the communication and cooperation between agents. [Guiraud 11] provides a BDI-based framework for emotion triggering and expression through expressive speech acts. Finally, FIPA provides a rich specification for intelligent agents communication that is based on speech acts [FIPA 02]. All this related work will help us define communication

and interaction rules in our model.

### 1.4.2 Sociality and Theory of Mind

Castelfranchi claims that social action cannot be reduced to communication. "[Agents] are not 'social' because they communicate, they communicate because they are 'social'" [Castelfranchi 97]. Indeed, according to him, sociality is defined by the way individuals act (i.e. cooperate, compete, organize,etc.) in a common world and interfere with, depend on, and influence each other. Consequently, goal delegation, goal adoption adoption, social manipulation, etc. form the basis of social interaction and collaboration. Therefore, given the key role of the theory of mind in this kind of interactions, modeling and implementing it necessary to create social agents.

[Castelfranchi 98] presents a theory of delegation for multi-agent systems. Although not fully formalized, it provides an interesting basis to model this kind of social behavior. [Herzig 02] also tackles principles such as *belief adoption* and *intention generation* based on assertive speech acts in their cooperation framework.

## 1.5 Discussion

In this chapter, we introduced the theoretical background needed to address and define our research topic. Additionally, we presented an overview of the related work in our discipline of interest, i.e. affective computing. The remaining of this document will describe our work as well as its evaluation.

# Chapter 2

# Logical framework

In this chapter, we will define the syntax and the semantics of our model. First, we will introduce the language of the logical framework. Then, we will present in more details the semantics of our model's operators that allow for representing mental states, social relations, emotions triggering and interaction mechanisms.

In the following, $\stackrel{\text{def}}{=}$ and $\stackrel{\text{def}}{\Longrightarrow}$ respectively mean *equals by definition* and *implies by definition*. The former is used to define new operators as functions of others and the latter to express rules such that when the premise is true, so is the conclusion.

## 2.1 Syntax

Assume a finite set of atomic propositions $ATM$, a finite set of physical actions $ACT$, a finite set of illocutionary (speech) acts $ILL$, a finite set of agents $AGT$, a finite set of emotions $EMO$, which is a subset of the twenty two OCC emotions, and the intervals of real numbers $DEG = [-1, 1]$ and $DEG^+ = [0, 1]$. $ATM$ describes facts or assertions such as *salary_is_bad* or *picnic_is_fun*. The actions $ACT$ that the agents may perform are expressed with verbs in the infinitive form, e.g. *introduce_itself* or *have_a_picnic*. $AGT$ includes animates, i.e. *Humans* and *Virtual agents*.

Our model defines events as acts in which at least one of the actors of the interaction take part. Contrariwise, events such as *rain_starts_falling* are represented by propositions in $ATM$. So, $EVT$ is formed by vectors of the following form : $\langle$*active agent, passive agent, content*$\rangle$. This representation is very similar to the one in [Ochs 09] except we do not include the *degree of certainty* in the vector. Indeed, we chose to represent a subjective – rather than objective – probability of a state of affair through a *degree of believability* as it will be explained later in this section. We consider two types of acts : actions and

speech acts.

As far as social relations are concerned, our model uses a two-dimensional representation. Thus, based on interpersonal theory models [Leary 57] [Kiesler 96], we model them according to the degree of *liking* and *dominance* an agent considers it has for and on another.

The language we define is the set of formulas described by the following BNF (Backus-Naur-Form):

$$\begin{aligned} Evt : \epsilon &::= \langle a, (a|\varnothing), \alpha \rangle \mid \langle a, a, Spk(\varsigma, \varphi) \rangle \\ Prop : \pi &::= p \mid \epsilon \mid Like^k_{a,b} \mid Dom^k_{a,b} \\ Fml : \varphi &::= \pi \mid Bel^l_a(\varphi) \mid Att^k_a(\varphi) \mid Int_a(\varphi) \mid Emo^i_{a,(b|\varnothing)}(\varepsilon, \varphi) \mid N(\varphi) \mid U(\varphi, \varphi) \mid \neg\varphi \mid \varphi \wedge \varphi \end{aligned} \tag{2.1}$$

where $a, b \in AGT$, $\alpha \in ACT$, $p \in ATM$, $\epsilon \in EVT$, $\varepsilon \in EMO$, $l, i \in DEG^+$, $k \in DEG$. *Emo* and *Spk* respectively describe speech acts and emotions, as it will be detailed in Section 2.2.5 and Section 2.2.6. *Bel*, *Att* and *Int* are modal operators and $N$, and $U$ are temporal operators. The other boolean conditions $\top$, $\bot$, $\vee$ and $\Rightarrow$ are defined in the standard way. Moreover, in the events' representation, $-$ is the *any* operator.

$Like^k_{a,b}$ determines the level of liking an agent has for another, i.e. its attitude towards it while $Dom^k_{a,b}$ represents the degree of dominance, control and/or power it has over it [Ochs 09]. These two relational factors are considered subjective and not necessarily symmetric.

$N$ and $U$ represent the standard temporal operators. $N(\varphi)$ means *"$\varphi$ will be true in the next iteration"* and $U(\varphi_1, \varphi_2)$ means *"$\varphi_1$ holds until $\varphi_2$ is true"*. We also introduce the other standard temporal operators $F$ and $G$ the way they are usually defined:

$$\begin{aligned} F(\varphi) &\stackrel{\text{def}}{=} U(\top, \varphi) \\ G(\varphi) &\stackrel{\text{def}}{=} \neg F(\neg\varphi) \end{aligned} \tag{2.2}$$

$Bel^l_a(\varphi)$ is a graded belief and has to be read *"a believes that $\varphi$ with certainty $l$"*. This is expressed in [Dastani 12] through an *exceptionality* operator, but in both cases, we consider plausibility is subjective, and represent it by the degree of believability of a formula for the agent, symbolized by $l$. This is why we do not define events' *degrees of certainty* like in [Ochs 09]. For instance, $Bel^1_a(\varphi)$ means *"a is sure that $\varphi$"* and $Bel^0_a(\varphi)$ can be read *"For a, $\varphi$ is not plausible at all"*.

Similarly, $Att^k_a(\varphi)$ is a graded attitude that has to be read *"a has a positive/negative attitude, with a degree of appreciation l, towards the fact that $\varphi$"* or simply *"a appreciates/values the fact that $\varphi$ with a degree $l$"*. In our context, we think this operator can

cover various notions, such as *Desires*, *Ideals* and *Goals*. Indeed, we believe the former can be seen as attitudes towards possible future states of affairs for instance. Moreover, unlike [Adam 09] and [Guiraud 11], we do not find it necessary to define a distinct operator to represent ideals, i.e. what is morally right or praiseworthy. We simply model this as *something an agent wants to be always true*. Finally, in our context of scenario-based Human/Agent interaction, although we agree that goals are chosen desires which have to be consistent and believed to be achievable as it is stated BDI theory [Rao 91], we do not express them separately. Hence, in our model, the subject of an attitude can as well be *preserving_forest*, *being_nice_to_others*, *hiring_new_employee* or $Bel^l_b(\langle a, c, give\_sandwich\rangle)$, eventually encapsulated in temporal operators.
Nevertheless, for the sake of readability, we define the graded desire operator $Des^k_a(\varphi)$ that can be read *"a wants $\varphi$ to be true with a degree of desirability k"* [Dastani 12] But, as explained above, we do it using the *Att* operator, as follows:

$$Des^k_a(\varphi) \stackrel{\text{def}}{=} Att^k_a(F(\varphi)) \tag{2.3}$$

$Int_a(\varphi)$ represents an agent's plan, something it commits to attempt to realize [Rao 91] and has to be read *"a intends to make $\varphi$ true"*.

As for $Emo^i_{a,(b|\varnothing)}(\varepsilon, \varphi)$, it has to be read *"a feels $\varepsilon$, eventually for/towards b, with intensity i, regarding the fact that $\varphi$"* with $\varepsilon \in EMO$. For the sake of simplification, in Section 2.2.5, we will write $\varepsilon^i_{a,(b|\varnothing)}(\varphi)$.

Likewise, $\langle a, b, Spk(\varsigma, \varphi)\rangle$ means *"a utters $\varphi$ to b by the illocutionary act $\varsigma$"* where $\varsigma \in ILL$ and will simply be written $\varsigma_{a,b}(\varphi)$ in Section 2.2.6.

For readability, we introduce new operators to represent agents' involvement in an event. $Resp_a$ expresses a *direct responsibility*, that is to say, unlike [Adam 12] and [Guiraud 11], we do not consider an agent responsible for a situation it could have avoided. $Wit_a$ means that the agent witnessed the occurrence of the event. As this model is only aimed for dyadic interaction, the are only two possible witnesses:

$$Resp_a(\epsilon) \stackrel{\text{def}}{=} (\epsilon = \langle a, -, -\rangle) \tag{2.4}$$

$$Wit_a(\epsilon) \stackrel{\text{def}}{=} (\epsilon = \langle a, -, -\rangle) \vee (\epsilon = \langle -, a, -\rangle) \tag{2.5}$$

## 2.2 Semantics

Based on possible world semantics, we define a frame $\mathcal{F} = \langle W, \mathcal{B}, \mathcal{D}, \mathcal{I}, \mathcal{E}\rangle$ as a tuple where:

- $W$ is a nonempty set of possible worlds,

- $\mathcal{B} : AGT \rightarrow (W \rightarrow 2^W)$ is the function that associates each agent $a \in AGT$ and possible world $w \in W$ to the set of belief-accessible worlds $\mathcal{B}_a(w)$,
- $\mathcal{D} : AGT \rightarrow (W \times DEG^+ \rightarrow 2^W)$ is the function that associates each agent $a \in AGT$ and possible world $w \in W$ with a level of desirability $l \in DEG^+$ to the set of desire-accessible worlds $\mathcal{D}_a(w, l)$,
- $\mathcal{I} : AGT \rightarrow (W \rightarrow 2^W)$ is the function that associates each agent $a \in AGT$ and possible world $w \in W$ to the set of intention-accessible worlds $\mathcal{I}_a(w)$, and
- $\mathcal{E} : EVT \rightarrow W$ is the function that associates each event $\epsilon \in EVT$ to the resulting possible world.

Then, a model $\mathcal{M} = \langle \mathcal{F}, \mathcal{V} \rangle$ is a couple where $\mathcal{F}$ is a frame and $\mathcal{V} : W \rightarrow ATM$ a valuation function.

Given a model $\mathcal{M}$ we note $\mathcal{M}, w \models \varphi$ a formula $\varphi$ that is true in a world $w$. Hence, we define truth conditions of formulas as follows:

- $\mathcal{M}, w \models p$ iff $p \in \mathcal{V}(w)$;
- $\mathcal{M}, w \models \neg\varphi$ iff not $\mathcal{M}, w \models \varphi$;
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$ ;
- $\mathcal{M}, w \models Bel^l_a(\varphi)$ iff $\frac{card(\mathcal{GB}_a(w))}{card(\mathcal{B}_a(w))} = l$ where $\mathcal{GB}_a(w) = \{v \in \mathcal{B}_a(w) \ ; \ \mathcal{M}, v \models \varphi\}$ ;
- $\mathcal{M}, w \models Des^l_a(\varphi)$ iff $\mathcal{M}, v \models \varphi \ \forall v \in \mathcal{D}_a(w, l)$;
- $\mathcal{M}, w \models Int_a(\varphi)$ iff $\mathcal{M}, v \models \varphi \ \forall v \in \mathcal{I}_a(w)$;
- $\mathcal{M}, w \models \epsilon$ iff $\mathcal{M}, v \models \top \ \forall v \in \mathcal{E}(\epsilon)$;

The truth condition of $Bel^l_a(\varphi)$ states that the level of plausibility of $\varphi$ equals the number of belief-accessible worlds where $\varphi$ is true divided by the the total number of possible worlds for agent $a$.

In the following sections, we define the semantics of the operators defined in the framework. Besides, please note that the *level functions* – indicating believability, desirability and intensity degrees in some reasoning and emotion triggering rules – will not be detailed in this chapter. We rather propose an implementation of these functions in the next chapter and leave open the possibility to readjust them in future work.

### 2.2.1 Graded beliefs

All accessibility relations $\mathcal{B}$ are transitive[1] and euclidean[2], which ensures that the agent is aware of its own beliefs[3]:

$$Bel_a^l(\varphi) \overset{\text{def}}{\Longrightarrow} Bel_a^1(Bel_a^l(\varphi)) \tag{2.6}$$

However, unlike other models [Adam 09] [Dastani 12], $\mathcal{B}$ is not serial[4]. Only $\mathcal{GB}$ is. Indeed, the agent generally has uncertainty about states of affairs. Intuitively:

$$Bel_a^l(\varphi) \overset{\text{def}}{\Longrightarrow} Bel_a^{1-l}(\neg\varphi) \tag{2.7}$$

For convenience, we define two thresholds[5] *mod_thld* and *str_thld* along with the operators $ModBel$, $StrBel$ and $SurBel$, respectively meaning *moderately*, *strongly* and *surely believes*, as following:

$$\begin{aligned} ModBel_a^l(\varphi) &\overset{\text{def}}{=} Bel_a^{mod\_thld<l<str\_thld}(\varphi) \\ StrBel_a^l(\varphi) &\overset{\text{def}}{=} Bel_a^{str\_thld\leq l<1}(\varphi) \\ SurBel_a^l(\varphi) &\overset{\text{def}}{=} Bel_a^1(\varphi) \end{aligned} \tag{2.8}$$

In the rest of this document, when the level of plausibility is not specified, by *"a believes that $\varphi$"* we implicitly mean *"a believes at least moderately that $\varphi$"*, that is to say *"a believes $\varphi$ is more likely than $\neg\varphi$"*:

$$Bel_a^l(\varphi) = ModBel_a^l(\varphi) \vee StrBel_a^l(\varphi) \vee SurBel_a^l(\varphi) = Bel_a^{l'>mod\_thld}(\varphi) \tag{2.9}$$

Furthermore, we generalize (2.6) so that agents are aware of their own mental states, social relations and involvement in events:

$$\begin{aligned} Att_a^k(\varphi) &\overset{\text{def}}{\Longrightarrow} SurBel_a^1(Att_a^k(\varphi)) \\ Int_a(\varphi) &\overset{\text{def}}{\Longrightarrow} SurBel_a^1(Int_a(\varphi)) \\ Like_{a,b}^k &\overset{\text{def}}{\Longrightarrow} SurBel_a^1(Like_{a,b}^k) \\ Dom_{a,b}^k &\overset{\text{def}}{\Longrightarrow} SurBel_a^1(Dom_{a,b}^k) \\ Resp_a(\epsilon) &\overset{\text{def}}{\Longrightarrow} SurBel_a^1(Resp_a(\epsilon)) \\ Wit_a(\epsilon) &\overset{\text{def}}{\Longrightarrow} SurBel_a^1(Wit_a(\epsilon)) \end{aligned} \tag{2.10}$$

[1] A given relation $\mathcal{R}$ is transitive iff if $w\mathcal{R}v$ and $v\mathcal{R}u$ then $w\mathcal{R}u$

[2] A given relation $\mathcal{R}$ is euclidean iff if $w\mathcal{R}v$ and $w\mathcal{R}u$ then $v\mathcal{R}u$

[3] If $w\mathcal{R}v$ and $v\mathcal{R}u$, then successively by transitivity, euclidianity and transitivity again: $w\mathcal{R}v$ and $v\mathcal{R}v$

[4] A given relation $\mathcal{R}$ is serial iff $\forall w$, $\exists v$ so that $w\mathcal{R}v$

[5] Thresholds *mod_thld* and *str_thld* are set to 0.5 and 0.75 in our implementation. If other values are to be chosen if the future, one must make sure that $0.5 < mod\_thld < str\_thld$

Finally, if an agent believes a state of affairs to possibly cause another, it will believe the latter with a proportional degree:

$$Bel_a^l(\psi) \wedge Bel_a^{l'}(\psi \Rightarrow \varphi) \stackrel{\text{def}}{\Longrightarrow} Bel_a^{f(l,l')}(\varphi) \tag{2.11}$$

### 2.2.2 Graded attitudes and desires

Attitudes towards states of affairs can be positive or negative. We assume that:

$$Att_a^{k\geq 0}(\varphi) \stackrel{\text{def}}{\Longrightarrow} Att_a^{-k\leq 0}(\neg\varphi) \tag{2.12}$$

However, an agent cannot hold inconsistent desires in the sense that:

$\mathcal{M}, w \models (Att_a^k(\varphi) \wedge Att_a^{k'}(\neg\varphi))$ iff $k \neq -k'$.

Subsequently, likewise, desires can be negative, which expresses an aversion for a state of affairs. For instance:

$Des_a^{k<0}(a\_gets\_sick) = Att_a^{k<0}(F(a\_gets\_sick)) = Att_a^{k>0}(G(\neg a\_gets\_sick))$
means *"a does not want to get sick"*. which does not express the same kind of undesirability than:

$Des_a^{k>0}(\neg a\_is\_jobless) = Att_a^{k>0}(F(\neg a\_is\_jobless)) = Att_a^{k<0}(G(a\_is\_jobless))$
which means that *"a wants b to stop being jobless"*, i.e. *"a does not want to be jobless forever"*. Please note that in both cases, for a $Des_a^k(\varphi)$ to be relevant, $\varphi$ should be currently false.

Although we excluded inconsistent desires in our definition, an *indirect* inconsistency is still possible: an agent might want something that can possibly lead to or be caused by (the occurrence of) the negation of another desire of his. Hence, in order to adopt a new desire, we must avoid this kind of inconsistency:

$$Des_a^k(\varphi) \wedge StrBel_a^l(\psi \Rightarrow F(\varphi)) \wedge \neg IncDes_a^k(\psi) \stackrel{\text{def}}{\Longrightarrow} N(Des_a^k(\psi)) \tag{2.13}$$

Where:

$$IncDes_a^k(\varphi) \stackrel{\text{def}}{=} (StrBel_a^l(\varphi \Rightarrow \neg\psi) \wedge Des_a^{k'>0}(\psi)) \vee (StrBel_a^l(\varphi \Rightarrow \psi) \wedge Des_a^{k'<0}(\psi)) \tag{2.14}$$

This means that desiring $\varphi$ is inconsistent when the agent strongly beliefs it might lead to an undesirable $\psi$. Thus, we still allow for adopting new indirectly inconsistent desires when the agent only believes *moderately* that there can be a certain incompatibility with existing ones.

One might have noticed from previous examples that desires having negative levels express long-term wills and attitudes, i.e. constant desires to maintain desirable states of affairs, such as *not getting sick*. They are not meant to generate *Intentions* in the sense that they do not imply any specific action performance but only avoiding – as much as possible – those that are incompatible. These are what we consider as *Ideals*:

$$Ideal_a^{k>0}(\varphi) \stackrel{\text{def}}{=} Att_a^{k>0}(G(\varphi)) = Des_a^{-k<0}(\neg\varphi) \tag{2.15}$$

On the other hand, desires having positive levels do serve for short-term objectives, those that have to produce an intention, and subsequently an act, in order to be fulfilled, like *passing an exam*. These are potential *Goals*.
In order to filter negative but also insufficiently strong desires, we define a new threshold[6] *des_thld*:

$$StrDes_a^k(\varphi) \stackrel{\text{def}}{=} Des_a^{k\geq des\_thld}(\varphi) \tag{2.16}$$

and a weaker case of inconsistency :

$$WIncDes_a^k(\varphi) \stackrel{\text{def}}{=} StrBel_a^l(\varphi \Rightarrow \neg\psi) \wedge Des_a^{k';|k'|>|k|}(\psi) \tag{2.17}$$

Where desiring $\varphi$ is considered inconsistent only if it leads to an undesirable state of affairs with a higher level. This way, according to the BDI model, we are able to define *Goals* as chosen desires that are consistent – at least weakly – and believed to be achievable [Rao 91]:

$$Goal_a^{k>0}(\varphi) \stackrel{\text{def}}{=} StrDes_a^k(\varphi) \wedge Bel_a^l(F(\varphi)) \wedge \neg WIncDes_a^k(\varphi) \tag{2.18}$$

Then, as for transitions between desires and intentions – through goals –, there are two cases. The first and simplest one is that of directly wanting to perform an act which is believed to be doable:

$$Goal_a^{k>0}(\epsilon) \wedge Resp_a(\epsilon) \stackrel{\text{def}}{\Longrightarrow} N(Int_a(\epsilon)) \tag{2.19}$$

Secondly, similarly to [Bosse 11], if the agent strongly believes there is – at least – one means to realize the selected desire, it will intend to perform it, provided that it is feasible, like suggested by [FIPA 02]:

$$Goal_a^{k>0}(\varphi) \wedge StrBel_a^l(\psi \Rightarrow F(\varphi)) \wedge \neg WIncDes_a^k(\psi) \wedge Bel_a^{l'}(F(\psi)) \stackrel{\text{def}}{\Longrightarrow} N(Int_a(\psi)) \tag{2.20}$$

We leave it to the implementation phase to decide how to order intentions when several ways to achieve a goal are known by the agent.

[6] threshold *des_thld* is set to 0.7 in our implementation

### 2.2.3 Intentions and acts

Since intentions are generated from desires, likewise, all accessibility relations $\mathcal{I}$ are serial. This can be formulated as follows:

$$Int_a(\varphi) \stackrel{\text{def}}{\Longrightarrow} \neg Int_a(\neg\varphi) \tag{2.21}$$

If an agent intends a state of affairs which it strongly believes to be eventually caused by another, it will also intends the latter:

$$Int_a(\varphi) \wedge StrBel^l_a(\psi \Rightarrow F(\varphi)) \stackrel{\text{def}}{\Longrightarrow} Int_a(\psi) \tag{2.22}$$

Additionally, if an agent intends an act which it is responsible for, it will perform it in the next step:

$$Int_a(\epsilon) \wedge Resp_a(\epsilon) \stackrel{\text{def}}{\Longrightarrow} N(\epsilon) \tag{2.23}$$

Furthermore, when an event occurs, we propagate responsibility to all the states of affairs it is believed to have caused:

$$Bel^d_a(\psi) \wedge Bel^l_a(Resp_b(\psi)) \wedge Bel^{l'}_a(\varphi) \wedge Bel^{l''}_a(\psi \Rightarrow F(\varphi)) \stackrel{\text{def}}{\Longrightarrow} Bel^{f(l,l',l'')}_a(Resp_b(\varphi)) \tag{2.24}$$

Finally, as far as accessibility relations $\mathcal{E}$ are concerned, we consider any witness know that an event happened and that the other knows that, too:

$$\epsilon \wedge Resp_a(\epsilon) \wedge Wit_b(\epsilon) \stackrel{\text{def}}{\Longrightarrow} G(SurBel^1_a(\epsilon)) \wedge G(SurBel^1_a(SurBel^1_b(\epsilon))) \tag{2.25}$$

Note that when an event occurs, the belief that it happened remains true afterwards.

### 2.2.4 Updating mental states

Beliefs are the informative component of the system [Rao 95] They are initialized in the interaction starting and then updated as new events occur; see (2.25). Anyway, the agent has to appraise current world states of affairs along with its held mental states in order to update the latter and react consequently.

[FIPA 02] suggests that, if an agent has a goal, it is committed to it until it is believed to be achieved or unachievable. Generalizing this principle to all the (positive) desires, we propose the following :

$$StrBel^l_a(\varphi) \wedge Des^{k>0}_a(\varphi) \stackrel{\text{def}}{\Longrightarrow} N(\neg Des^k_a(\varphi)) \wedge N(\neg Int_a(\varphi)) \tag{2.26}$$

$$StrBel^l_a(\neg F(\varphi)) \wedge Des^{k>0}_a(\varphi) \stackrel{\text{def}}{\Longrightarrow} N(\neg Des^k_a(\varphi)) \wedge N(\neg Int_a(\varphi)) \tag{2.27}$$

Nevertheless, ideals are supposed to be constant and to hold globally.

In order for the agent to be able to react to current world state, attitudes about new states of affairs have to be triggered. We argue that the way one appraises a new state of affairs depends on the context, which, in our case, consists of the dyadic interaction. Therefore, we propose that the agent's attitude depends also on the attributed other's and on the social relation between them:

$$StrBel_a^l(\varphi) \wedge Att_a^k(F(\varphi)) \wedge Bel_a^{l'}(Att_b^{k'}(F(\varphi))) \wedge Like_{a,b}^h \wedge Dom_{a,b}^{h'} \stackrel{\text{def}}{\Longrightarrow} Att_a^{f(k,k',h,h')}(\varphi) \quad (2.28)$$

### 2.2.5 Emotions triggering

In this section, we will define the set of emotions $EMO$ based on their triggering conditions as presented in the OCC theory [Ortony 90]. However, we will not model the *Attraction* emotion, i.e. *Love* and *Hate*, since they are directed towards individuals rather than states of affairs and we do not find this relevant in our context. We will first introduce the factors that can influence the emotions' intensity. Then, as in [Adam 09] and [Dastani 12], we will define the triggering conditions based on an evaluation of states of affairs, that in our case include events, propositions and formulas.

We can divide emotions in two groups: *directed* and *non-directed* emotions. The arity of the former is 4 and the latter's is 3 (agents involved + emotion's intensity + subject). Even though reflexive emotions, such as *Pride*, are directed, their formalizations only take two arguments and thus are associated to the second group.

#### Factor influencing emotions' intensity

In our model, the intensity of an emotion depends on the plausibility and the desirability degrees of the triggering states of affairs and on the attitude towards the passive agent (in the case of directed emotions), which is consistent with appraisal theories [Ortony 90][Lazarus 91]. Although not as exhaustive, these three variables allow us to model a satisfying number of the factors enumerated in OCC theory.

Indeed, the degree of certainty of graded beliefs can be used to illustrate 1) the *sense of reality* which "depends on how much one believes the emotion-inducing situation is real", 2) the *unexpectedness* which "depends on how suprised one is by the situation", 3) the *likelihood* which "reflects the degree of belief that an anticipated event will occur" and 4) the *realization* which "depends on the degree to which an anticipated event actually occurs" [Ortony 90]. These factors can also be calculated for others, as long as the Theory of mind model generates attributed mental states for them (see Section 3.1).

Similarly, the *Desire* operator covers both *desirability-for-self* and presumed *desirability-*

*for-others* generated by ToM. Besides, as explained previously, *praiseworthiness* is represented by *Ideals* (see Section 2.2).

Finally, in OCC theory, the fortune-of-others emotions are influenced by the attitude an agent has for another, i.e. the *liking* level, [Ortony 90]. Nevertheless, since social relations affect the attitude towards current states of affairs (see (2.28)), they indirectly influence others categories of emotions as well.

For convenience, let $\mathscr{X}(\varphi)$ symbolize a formula that does not imply any temporal operator, i.e. that is not of the form $N(\varphi)$ or $U(\varphi)$.

In all the following triggering rule, let $\gamma = \mathscr{X}(\varphi)$.

### Well-being emotions

[Ortony 90] suggests that these emotions are "essentially 'pure' cases of being pleased or displeased" and that the main factor affecting their intensity is the degree of desirability of the event. However, in our interpretation, the level of certainty also influences it. This allows us to define the following triggering rules:

$$\begin{aligned} &Bel_a^l(\gamma) \wedge Att_a^{k>0}(\gamma) \overset{\text{def}}{\Longrightarrow} N(Joy_a^{i=f(l,k)}(\gamma)) \\ &Bel_a^l(\gamma) \wedge Att_a^{k<0}(\gamma) \overset{\text{def}}{\Longrightarrow} N(Distress_a^{i=f(l,k)}(\gamma)) \end{aligned} \tag{2.29}$$

### Prospect-based emotions

This class is based on the ability to expect the occurrence of an event. The first pair of emotions of this category is similar to the well-being emotions except, here, the agent would appraise an eventual state of affairs. To express the anticipation process, we use the temporal operator $Evn$. The intensity of the triggered emotions then depends on the desirability of a state of affairs and on the likelihood of its occurrence [Ortony 90]:

$$\begin{aligned} &Bel_a^l(F(\gamma)) \wedge Des_a^{k>0}(\gamma) \overset{\text{def}}{\Longrightarrow} N(Hope_a^{i=f(l,k)}(\gamma)) \\ &Bel_a^l(F(\gamma)) \wedge Des_a^{k<0}(\gamma) \overset{\text{def}}{\Longrightarrow} N(Fear_a^{i=f(l,k)}(\gamma)) \end{aligned} \tag{2.30}$$

Subsequently, depending on whether the anticipated event happens, another group of emotions can be triggered. Their intensity will be influenced by the underlying hope's (or fear's) strength, which, as mentioned above, we calculate in terms of likelihood and

desirability, and by the level of certainty the agent has on its actual occurrence.

$$\begin{array}{rcl} Hope_a^{i=f(l,k)}(\gamma) \wedge Bel_a^d(\gamma) & \stackrel{\text{def}}{\Longrightarrow} & N(Satisfaction_a^{i=f(l,k,d)}(\gamma)) \\ Hope_a^{i=f(l,k)}(\gamma) \wedge Bel_a^d(\neg\gamma) & \stackrel{\text{def}}{\Longrightarrow} & N(Disappointment_a^{i=f(l,k,d)}(\neg\gamma)) \\ Fear_a^{i=f(l,k)}(\gamma) \wedge Bel_a^d(\gamma) & \stackrel{\text{def}}{\Longrightarrow} & N(FearConfirmed_a^{i=f(l,k,d)}(\gamma)) \\ Fear_a^{i=f(l,k)}(\gamma) \wedge Bel_a^d(\neg\gamma) & \stackrel{\text{def}}{\Longrightarrow} & N(Relief_a^{i=f(l,k,d)}(\neg\gamma)) \end{array} \tag{2.31}$$

**Fortune-of-others emotions**

This class of emotions highly depends on the mental states one imputes to others and is hence strongly connected to the Theory of Mind. However, the aim of this section is not to discuss the way those attributed mental states are generated (e.g. by commonsense rules or by a mirroring process). Here, we just suppose they are held in the agent's knowledge base, i.e. its set of beliefs (see Section 3.1).

When one is pleased or displeased by the occurrence of events, depending on whether they are presumed to be desirable or undesirable for another agent and on the relation between them, the following emotions can be triggered:

$$\begin{array}{rcl} Bel_a^d(\gamma) \wedge Bel_a^l(Att_b^{k>0}(\gamma)) \wedge Like_{a,b}^{k'>0} & \stackrel{\text{def}}{\Longrightarrow} & N(HappyFor_{a,b}^{i=f(l,k,k',d)}(\gamma)) \\ Bel_a^d(\gamma) \wedge Bel_a^l(Att_b^{k<0}(\gamma)) \wedge Like_{a,b}^{k'>0} & \stackrel{\text{def}}{\Longrightarrow} & N(SorryFor_{a,b}^{i=f(l,k,k',d)}(\gamma)) \\ Bel_a^d(\gamma) \wedge Bel_a^l(Att_b^{k>0}(\gamma)) \wedge Like_{a,b}^{k'<0} & \stackrel{\text{def}}{\Longrightarrow} & N(Resentment_{a,b}^{i=f(l,k,k',d)}(\gamma)) \\ Bel_a^d(\gamma) \wedge Bel_a^l(Att_b^{k<0}(\gamma)) \wedge Like_{a,b}^{k'<0} & \stackrel{\text{def}}{\Longrightarrow} & N(Gloating_{a,b}^{i=f(l,k,k',d)}(\gamma)) \end{array} \tag{2.32}$$

Notice that these rules can as well trigger, in the way it is suggested by intuition, the right emotions when events do not occur. For instance, one can be happy because of the non-occurrence of an event it believes to be undesirable for an agent it likes.

**Attribution emotions**

In order to distinguish this class of emotions from the previously described event-based ones, we rely on the *responsibility* operator. Indeed, as expressed in [Ortony 90], here "we focus on an agent whom we take to have been instrumental in the event, rather than the event itself". The triggering conditions are the following:

$$\begin{array}{rcl} Bel_a^l(\gamma) \wedge Ideal_a^k(\gamma) \wedge Bel_a^{l'}(Rsp_a(\gamma)) & \stackrel{\text{def}}{\Longrightarrow} & N(Pride_a^{i=f(l,l',k)}(\gamma)) \\ Bel_a^l(\gamma) \wedge Ideal_a^k(\neg\gamma) \wedge Bel_a^{l'}(Rsp_a(\gamma)) & \stackrel{\text{def}}{\Longrightarrow} & N(Shame_a^{i=f(l,l',k)}(\gamma)) \\ Bel_a^l(\gamma) \wedge Ideal_a^k(\gamma) \wedge Bel_a^{l'}(Rsp_b(\gamma)) & \stackrel{\text{def}}{\Longrightarrow} & N(Admiration_{a,b}^{i=f(l,l',k)}(\gamma)) \\ Bel_a^l(\gamma) \wedge Ideal_a^k(\neg\gamma) \wedge Bel_a^{l'}(Rsp_b(\gamma)) & \stackrel{\text{def}}{\Longrightarrow} & N(Reproach_{a,b}^{i=f(l,l',k)}(\gamma)) \end{array} \tag{2.33}$$

#### Compound emotions

*Gratification*, *Remorse*, *Gratitude* and *Anger* are defined in [Ortony 90] as Well-being/Attribution emotions, triggered when one both focuses on the praiseworthiness of an action and on its desirability. However, in our model, ideals are extracted from attitudes and, although with (2.28) the attitude responsible for the Well-being part differs from that representing praiseworthiness, we do not find it relevant to define compound emotions this way.

Nevertheless, similarly to [Guiraud 11] we think that one might distinguish *Gratitude* and *Anger* from *Admiration* and *Reproach* if the triggering state of affairs corresponds to a goal, that is to say it is not only praiseworthy but is also desirable and consistent enough to generate an intention of achievement:

$$\begin{aligned} Bel_a^l(\gamma) \wedge Ideal_a^k(\gamma) \wedge Bel_a^{l'}(Rsp_b(\gamma)) \wedge Goal_a^{k'}(\gamma) &\overset{\text{def}}{\Longrightarrow} N(Gratitude_{a,b}^{i=f(l,l',k,k')}(\gamma)) \\ Bel_a^l(\gamma) \wedge Ideal_a^k(\neg\gamma) \wedge Bel_a^{l'}(Rsp_b(\gamma)) \wedge Goal_a^{k'}(\neg\gamma) &\overset{\text{def}}{\Longrightarrow} N(Anger_{a,b}^{i=f(l,l',k)}(\gamma)) \end{aligned} \tag{2.34}$$

### 2.2.6 Speech acts and social interaction modeling

Since we are not interested in linguistic semantics and in analyzing the surface meaning of an utterance, we only represent the illocutionary and perlocutionary acts in our model, i.e. what is respectively done *in* and *by* the utterance.

#### Illocutionary acts

Searle distinguishes five kinds of illocutions: Assertives, Directives, Commissives, Expressives and Declarations. However, the latter are not so relevant in the sort of interaction we need to model and can be considered as part of the first category. Therefore, $ILL = \{Assert, Request, Commit, Express\}$. According to [Austin 62] [Searle 69] and [Davis 79], one distinction between illocutionary and perlocutionary acts is that the former are conventional while the latter are not or at least not necessarily. Based on what we consider as the normal intended effects in generic cases and on those among Searle's *rules for the use of illocutionary forces* [Searle 69] that are relevant given our sematics, we

define the *normal* speech acts triggering conditions as follows:

$$\begin{aligned}
\neg SurBel_a^1(SurBel_b^1(\varphi)) \wedge Int_a(StrBel_b^l(StrBel_a^{l'}(\varphi))) &\stackrel{\text{def}}{\Longrightarrow} Assert_{a,b}(\varphi) \\
\neg SurBel_a^1(Int_b(\varphi)) \wedge Int_a(Int_b(\varphi)) &\stackrel{\text{def}}{\Longrightarrow} Request_{a,b}(\varphi) \\
\neg SurBel_a^1(SurBel_b^1(Int_a(\varphi))) \wedge Int_a(StrBel_b^l(Int_a(\varphi))) &\stackrel{\text{def}}{\Longrightarrow} Commit_{a,b}(\varphi) \\
\neg SurBel_a^1(SurBel_b^1(\varepsilon^i_{a,(b|\varnothing)}(\varphi))) \wedge Int_a(StrBel_b^l(\varepsilon^i_{a,(b|\varnothing)}(\varphi))) &\stackrel{\text{def}}{\Longrightarrow} Express_{a,b}(\varepsilon^i_{a,(b|\varnothing)}(\varphi))
\end{aligned} \tag{2.35}$$

Please note that this is consistent with Searle's suggested necessity to "capture both the intentional and the conventional aspect" of an illocution and that our definition do not cover non-particular behaviors such as sarcasm.

Assuming that the interlocutor correctly hears (receives) the messages but also speaks the same language and thus understands the surface sense of the utterance, a special case of the rule (2.25) is that a direct consequence of a speech act performance, is that the witnesses will surely believe that the event actually happened (see Section 2.2). Then, according to Searle, the illocutionary effect consists of the hearer's recognition that the states of affairs specified by (some of) the rules and convention of the common language obtain [Searle 69]. Here, we make – or at least make the agent make – the assumption of mutual belief in a sense that the hearer shares the same conventions and gets the meaning of its illocutions:

$$\begin{aligned}
Assert_{a,b}(\varphi) &\stackrel{\text{def}}{\Longrightarrow} StrBel_b^d(\neg SurBel_a^1(SurBel_b^1(\varphi)) \wedge Int_a(StrBel_b^l(StrBel_a^{l'}(\varphi)))) \\
Request_{a,b}(\varphi) &\stackrel{\text{def}}{\Longrightarrow} StrBel_b^d(\neg SurBel_a^1(Int_b(\varphi)) \wedge Int_a(Int_b(\varphi)) \\
Commit_{a,b}(\varphi) &\stackrel{\text{def}}{\Longrightarrow} StrBel_b^d(\neg SurBel_a^1(SurBel_b^1(Int_a(\varphi))) \wedge Int_a(StrBel_a^l(Int_a(\varphi)))) \\
Express_{a,b}(\varepsilon^i_{a,(b|\varnothing)}(\varphi)) &\stackrel{\text{def}}{\Longrightarrow} StrBel_b^d(\neg SurBel_a^1(SurBel_b^1(\varepsilon_a)) \wedge Int_a(StrBel_a^l(\varepsilon_a)))
\end{aligned} \tag{2.36}$$

While (2.35) aims to generate a richer behavior and allow for interactions in a non-scenario-based way, (2.36) would increase the number of attributed mental states and enrich the triggered emotion and eventually coping reactions.

### Perlocutions and social interaction

Regarding perlocutionary effects, as pointed out by [Davis 79] and [Marcu 00] the actual results of speech acts depend on various factors suchs as the speaker's and the audience's mental states, their relation, etc. In this framework, we model some of the social interaction as perlocutions resulting from speech acts. Please note that the following rules are relevant with several elements in the delegation theory presented in [Castelfranchi 98], as some conditions have been expressed in the triggering rules of the illocutions.

**Relation's influence on credibility** Whether an agent believes what another says depends on their relation:

$$Assert_{b,a}(\varphi) \wedge Like^{k}_{a,b} \wedge Dom^{k'}_{a,b} \overset{\text{def}}{\Longrightarrow} N(Bel^{f(k,k')}_{a}(\varphi)) \tag{2.37}$$

**Submission and obligation** A request from an other agent to which it is submissive will cause the agent to intend to do what has been asked:

$$Request_{b,a}(\varphi) \wedge Dom^{k<0}_{a,b} \overset{\text{def}}{\Longrightarrow} N(Int_{a}(\varphi))) \tag{2.38}$$

**Empathy and desire adoption** A desire expressed by an other agent it likes will cause the agent to adopt it:

$$StrBel^{l}_{a}(Des^{k}_{b}(\varphi)) \wedge Like^{k'>0}_{a,b} \overset{\text{def}}{\Longrightarrow} N(Des^{f(k,k')}_{a}(\varphi))) \tag{2.39}$$

## 2.3 Examples

### 2.3.1 Example 1: Lucy in the Forest with Mushrooms

Consider Lucy walking with her friend in the forest where they are going to have a picnic. Before she left home, Lucy's mother warned her of the possibility of getting sick if she eats an unknown mushroom, which they both want to avoid. This can be written this way:

| | |
|---|---|
| $Ideal^{0.8}_{Lucy}(\neg Lucy\_gets\_sick)$ | input |
| $\Longrightarrow Des^{-0.8}_{Lucy}(Lucy\_gets\_sick)$ | See (2.15) |
| $\Longrightarrow Att^{-0.8}_{Lucy}(F(Lucy\_gets\_sick))$ | See (2.3) |
| $\Longrightarrow Att^{0.8}_{Lucy}(\neg F(Lucy\_gets\_sick))$ | See (2.12) |
| $Ideal^{0.8}_{LucyMum}(\neg Lucy\_gets\_sick)$ | input |
| $Bel^{0.6}_{Lucy}(F(\epsilon_{LucyEUM}))$ where $\epsilon_{LucyEUM} = \langle Lucy, -, eat\_unknown\_mushroom\rangle$ | input |
| $\Longrightarrow Resp_{Lucy}(\epsilon_{LucyEUM})$ | See (2.4) |
| $StrBel^{0.85}_{Lucy}(\epsilon_{LucyEUM} \Longrightarrow F(Lucy\_gets\_sick))$ | input |

Now suppose Lucy indeed sees an unknown mushroom and is quite tempted – and, consequently, has a new strong desire – to try it. This leads to a weak indirect inconsistency with her other desires and ideals and keeps her from adopting it as a goal and doing it:

| | |
|---|---|
| $Des^{0.7}_{Lucy}(\epsilon_{LucyEUM})$ | input |
| $\Longrightarrow StrDes^{0.7}_{Lucy}(\epsilon_{LucyEUM})$ | See (2.16) |
| $\Longrightarrow WIncDes^{0.7}_{Lucy}(\epsilon_{LucyEUM})$ | See (2.17) |

Nevertheless, if Lucy is tempted enough by the mushroom, she can as well intend to taste it:

| | |
|---|---|
| $Des^{0.9}_{Lucy}(\epsilon_{LucyEUM})$ | input |
| $\Longrightarrow StrDes^{0.9}_{Lucy}(\epsilon_{LucyEUM})$ | See (2.16) |
| $\Longrightarrow \neg WIncDes^{0.9}_{Lucy}(\epsilon_{LucyEUM})$ | See (2.17) |
| $\Longrightarrow Goal^{0.9}_{Lucy}(\epsilon_{LucyEUM})$ | See (2.18) |
| $\Longrightarrow Int_{Lucy}(\epsilon_{LucyEUM})$ | See (2.19) |
| $\Longrightarrow \epsilon_{LucyEUM}$ | See (2.23) |
| $\Longrightarrow SurBel^{1}_{Lucy}(\epsilon_{LucyEUM})$ | See (2.25) |

Consequently, if Lucy actually gets sick after she ate the mushroom, she will feel distressed and ashamed about it:

| | |
|---|---|
| $SurBel^{1}_{Lucy}(Lucy\_gets\_sick)$ | input |
| $\Longrightarrow Att^{?<0}_{Lucy}(Lucy\_gets\_sick)$ | See (2.28) |
| $\Longrightarrow Distress^{?}_{Lucy}(Lucy\_gets\_sick)$ | See (2.29) |
| And | |
| $\Longrightarrow Bel^{?}_{Lucy}(Resp_{Lucy}(Lucy\_gets\_sick))$ | See (2.24) |
| $\Longrightarrow Shame^{?}_{Lucy}(Lucy\_gets\_sick)$ | See (2.33) |

### 2.3.2 Example 2: Gone Daddy's Gone

Consider two friends John and Mary having a conversation about their holidays. Mary is going to her home town. The fact that she is going to visit her father is a detail she could either mention or not:

| | |
|---|---|
| $Des^{0.77}_{Mary}(talking\_about\_holidays)$ | input |
| $StrBel^{0.8}_{Mary}(\langle Mary, John, visiting\_hometown\_and\_dad\rangle \Longrightarrow F(talk\_about\_holidays))$ | input |
| $StrBel^{0.8}_{Mary}(\langle Mary, John, visiting\_hometown\rangle \Longrightarrow F(talk\_about\_holidays))$ | input |

Nevertheless Mary remembers John recently lost his father and thus supposes it is a sensitive topic:

| | |
|---|---|
| $Bel^{1}_{Mary}(John\_lost\_his\_dad)$ | input |
| $StrBel^{0.76}_{Mary}(John\_lost\_his\_dad \Longrightarrow Ideal^{0.8}_{John}(F(\langle -, John, dad\rangle)))$ | input |
| $\Longrightarrow StrBel^{l}_{Mary}(Ideal^{0.8}_{John}(F(\langle -, John, dad\rangle)))$ | See (2.11) |

Of course, Mary knows that saying she is going to visit her father implies actually talking about her father:

$$StrBel^{0.8}_{Mary}(\langle Mary, -, visiting\_hometown\_and\_dad\rangle \Longrightarrow \langle Mary, -, dad\rangle) \qquad \text{input}$$

And, knowing that John wants to avoid this topic, she does too. Hence, she is will not mention the fact that she is visiting her father when talking about her holidays:

$$\begin{aligned}
&\Longrightarrow Ideal^{k}_{Mary}(F(\langle -, John, dad\rangle)) && \text{See (2.39)}\\
&\Longrightarrow WIncDes^{0.77}_{Mary}(\langle Mary, John, visiting_hometown_and_dad\rangle) && \text{See (2.17)}\\
&\Longrightarrow Goal^{0.77}_{Mary}(\langle Mary, John, visiting_hometown\rangle) && \text{See (2.18)}
\end{aligned}$$

### 2.3.3 Example 3: All apologies...

Consider James telling his friend Ana he lost her favorite book:

$$\begin{aligned}
&Bel^{0.8}_{James}(Ideal^{0.8}_{Ana}(\neg Lost_favorite_book)) && \text{input}\\
&Assert_{James,Ana}(\langle James, -, Lost_favorite_book\rangle) && \text{input}\\
&\quad\Longrightarrow Bel^{l}_{Ana}(Lost_favorite_book) && \text{See (2.37)}
\end{aligned}$$

By simulation-based mindreading, James can see she reproaches him for that (see (2.33)). But maybe she could forgive him if he apologizes...

## 2.4 Discussion

Based on the theoretical background we presented in the former chapter, we designed a logical framework that allows the agent to reason about others' mental and emotional states as well as to communicate and cooperate with them in the context of an interaction. The following phase of our work involves the definition and the implementation of a general architecture for a computational module relying on it.

# Chapter 3

# Module architecture and implementation

Previously, we presented a logical framework for representing the agent's mental states, emotions, actions and social relations and interactions. In Section 3.1, we introduce the general architecture that will allow us to model the agent's reasoning and theory of mind. Besides, the remaining section of this chapter describe the technological choices we made as well as some details regarding the implemention of this model and its connection with the TARDIS project.

## 3.1 Reasoning architecture

The module's general architecture is illustrated in Figure 3.1 and includes two main components:

**Agent's mental states** that encompass its Beliefs, Attitudes and Intentions:

- **Beliefs**: Agent's beliefs represent the informative aspect of the architecture and thus all the knowledge it can have. First, it has beliefs about its own mental states like stated in (2.6) and (2.10). Then, it has beliefs about others' mental states, i.e. attributed mental states, that can be acquired through speech acts or by commonsense reasoning. It is also aware of its world's current states of affairs. Finally, it knows some facts and rules about the functionning of its world, i.e. its notion – maybe subjective – of commonsense.
- **Attitudes**: Besides its attitudes about current states of affairs, which are *linking*

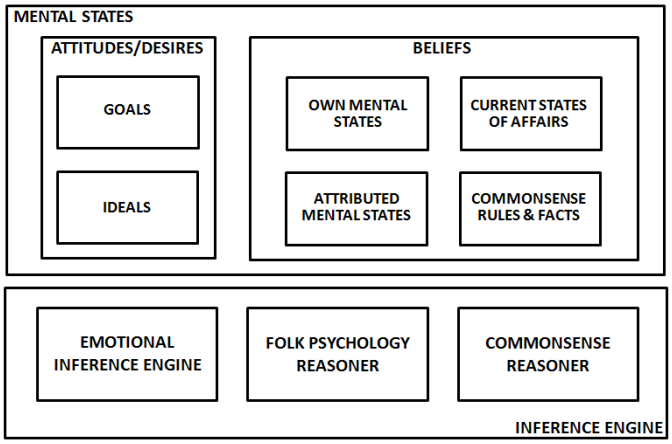



Figure 3.1: General architecture

attitudes, the agents holds has desires and ideals, i.e. respectively states of affairs its wants to occurs in the future or to be always true.

- **Intentions**: They express the deliberative aspect and are selected from agent's goals. Intentions are not represented in Figure 3.1 because in the first implementation of the model we are addressing in this chapter they are not fully part of the reasoning process but rather only giving as an output for action selection.

**Agent's inference engine** that comprises 3 modules:

- **Emotional inference engine**: This module is based on OCC-like rules allowing the agent to appraise its world's states of affairs and triggering the corresponding emotions according to its mental states.
- **Folk-psychology reasoner**: This is a deliberative reasoner that allows for intention generation according to the agent's beliefs and attitudes (Desires, Goals and Ideals). It is also responsible for updating its mental states.
- **Commonsense reasoner**: This additional module lets the agents deduce new beliefs – through the commonsense rules and facts – that can be used in the action selection process.

### 3.1.1 ToM TT and ST modeling

As we mentioned in Section 1.1.2, a TT approach for mindreading is based on the use of folk-psychology and/or commonsense to reason about the others while a ST approach

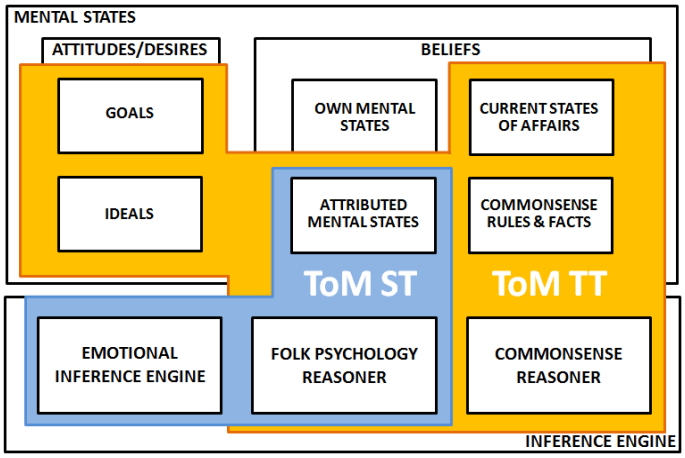



Figure 3.2: Modules involved in TT and ST ToM modeling

would suggest to project their attributed mental states on the agent's own inference engine. Consequently, Figure 3.2 shows what connections between the module presented previously allow for modeling these two theories of ToM within our architecture.

## 3.2 Technological choices

Prolog is a logic programming language based on first-order logic. It is declarative, which means that the program enumerates a set a predicates defining facts and rules and it is the compiler that transforms it into a sequence of instructions. Then, the computation consists of running a query and checking whether the goal clause can be proven. This is done by constructing a search tree and using the SLD (Selective Linear Definite clause) resolution method [The art of prolog]. Therefore, Prolog appears to be a suitable choice in order to implement the rules of our logical framework into an inference engine. Moreover, the resolution algorithm makes it possible to browse all the knowledge base and look for all the possibilities with an acceptable computational cost.

However, the inference engine needs to be integrated in other programs for it to be used by virtual conversational agents. As explained in the Introduction of this document, we are interested in plugging it to TARDIS and MARC projects. The former is mainly development in C++ language with the latter is in Java. Nevertheless, the priority is given to the connection with TARDIS. Besides, we do not need a fully object-oriented paradigm in the context of this project. Hence, we choose C++ to develop the module in charge of the interfacing with the environment and the launching of the goals to be proven by the Prolog reasoning engine. In this project, we used the MinGW 4.6.2 version of g++ compiler.

SWI-prolog offers a free Prolog environment. Additionally, it provides a powerful and flexible API that allows for embedding its kernel in C++ programs as well as a linker to generate an executable that combines all the Prolog and C++ files. Additionally, for future work, an embedding in Java programs is also possible. In this project, we used the 6.2.6 version of SWI-Prolog.

## 3.3 Reasoning engine implementation details

The architecture introduced in Section 3.1 defines the decomposition we adopted in order to encapsulate the rules presented in the logical model in different modules and, thereby, to organize the agent's reasoning process. This modularity has been kept in the implementation. Indeed, Prolog allows for creating modules as sets of predicates, some of which can be private, i.e. hidden, while others define the public interface being usable by other modules. This makes the code more readable and the control of the dependencies easier in the case of large programs. In addition, it makes the transposition of our theoretical architecture quite straightforward.

**The Action selection** module contains the set of rules defined in the logical model that allows for a BDI-like reasoning aiming to generate the agent's actions according to the current states of affairs.

**The Commonsense** module holds additional rules that may be domain-specific or not and help enrich the agent's behavior by providing more beliefs about the possible worlds and how it can satisfy its desires.

**The Emotion triggering** module lists the rules defining the OCC-like appraisal modeling based on its own and others' mental states.

**The facts base** contains the agent's own mental states and the attributed mental states, i.e. its beliefs about others' mental states, both provided at the program's initialization and acquired or updated during the interaction.

One of the difficulties we encountered during this phase is the implementation of some modal operators, especially the temporal ones. Indeed, Prolog is based on first-order logic and handling the temporal aspect as well as the equivalences induced by different combinations of temporal and other logical operators (e.g. $\neg$ and $\Rightarrow$) is not trivial.

Besides, as suggested in [Rao 95], our model cannot simply be implemented by a theorem-proving system, even if the temporal and epistemic aspects are handled. This is due to the fact that the computational cost for this kind of reasoning might be too important and thus affect the agent reactivity. For that matter, let us point out that Prolog is *not* a full logic programming language. Beside the declarative aspect, there is a procedural aspect to be taken into account, such as the fact that clauses are tested from top to bottom and their elements read from left to right.

For instance, this makes it quite complicated to write a clause that states a simple equivalence, determined by (2.7) and the classic temporal logic semantics, such as the following:

$$Bel_a^l(F(\varphi)) \stackrel{\text{def}}{=} Bel_a^{1-l}(\neg F(\varphi)) \stackrel{\text{def}}{=} Bel_a^{1-l}(G(\neg\varphi)) \stackrel{\text{def}}{=} Bel_a^l(\neg G(\neg\varphi))$$

However, the sequential aspect of Prolog programming can also be used to reduce the computational cost. Thus, instead of having an equivalence clause that the reasoner would have to call several times, we can assert the equivalent beliefs as new ones and add them to the database once and for all, right before any reasoning process is run. Please refer to Section 3.3.1 for more details about the reasoning loop.

Moreover, in rules (2.18) and (2.20) that allow for intention triggering for example, checking all the desires that might cause an inconsistency can be costly. Therefore, when implementing this process, we create an ordered goals list from desires and only generate a new intention if it is not inconsistent with an existing one. The condition on desirability degree is then implicitly verified.

### 3.3.1 Reasoning loop

In [Rao 95], a BDI interpreter abstract architecture is proposed as more *practical perspective* of the formal model presented in [Rao 91]. During every cycle, the agent would interpret external events to generate a list of potential actions, deliberate to select one of them, update its intentions and then execute them. In our module, the reasoning loop is quite similar except that intentions are executed in the very beginning.

In the theoretical model, we designed some behavioral and affective rules using the $N$ (Next) operator to illustrate the triggering aspect. Thus, the effects would not be instantaneous in case the temporal pace of the system is too short. However, in the context of implementation we are considering here, i.e. Question/Answer interaction, this delay might reduce the agent reactivity. Therefore, this operator has not been implemented and the delay has been discarded in most of the cases. For instance, during the deliberative process, we can see in (2.19) and (2.20) that an eligible goal generates an intention in the next iteration, and then in (2.23) that an intended event (or act) will be true in the

following, which will allow the agent to integrate its occurrence in its upcoming reasoning through (2.25). Removing the delays and making the agent execute its intention in the beginning of the following iteration allows for a compromise between too much and not enough reactivity.

The reasoning loop is the following:

$$Loop \begin{cases} Execute\_intentions \\ 1)\, Deduce\_by\_commonsense \\ 2)\, Simulate\_others\_emotions \\ 3)\, Update\_beliefs\_and\_attitudes \begin{cases} 3.1)\, Update\_beliefs\_with\_new\_SoA \\ 3.2)\, Handle\_operators\_equivalences \\ 3.3)\, Adopt\_new\_desires \\ 3.4)\, Order\_goals \end{cases} \\ 4)\, Adopt\_new\_intentions \begin{cases} 4.1)\, Adopt\_new\_intentions\_from\_goals \\ 4.2)\, Adopt\_new\_intentions\_from\_intentions \end{cases} \end{cases}$$

### 3.3.2 Job interview implementation

The course of the interview is handled in the commonsense module. In order to do the interview, the agent *believes* it has to go through six main parts, each one consisting in a list of topics it can address using speech acts. For example, the last part of the interview involves topics such as the salary or the schedules and practical questions like the earliest availability date. The agent also has *expectations* about the affective impact of the speech acts, which allows it to choose to avoid them or not according to its *goals*. Moreover, the agent can evaluate the candidates on three criteria: self-confidence, motivation and qualification. Indeed, it has a set of *rules* about how to interpret their affective reactions depending on the ongoing topic. For instance, a hesitation in the job description topic can indicate they are not qualified enough while being focused when introducing themselves denotes a good self-confidence level.

An interesting behavior emerged from the way we handle the interview progression. Indeed, when all the topics of the current part have been addressed, the reasoning process requires an additional iteration to be able to perform the first speech act of the following one. This generates some silences in the interaction that help structuring it and make it seem more realistic than a "mechanical" series of questions and answers.

### 3.3.3 Level functions implementation

When defining our logical framework's semantics in Chapter 2, we introduced some *level functions* allowing the calculation of the degrees of believability and desirability of new

mental state or the intensity of new emotions when they are triggered by logical rules. Those functions remained undefined because we believe their implementation may depend on the system's context of use and scope of application. Nevertheless, in this section, we will explain our approach regarding them when implementing the theoretical model and give some detailed examples.

All the rules in Section 2.2.5 rely on such functions to determine the intensity of the triggered emotions. Let us examine the example of *Joy* defines in (2.29):

$$Bel_a^l(\gamma) \wedge Att_a^{k>0}(\gamma) \stackrel{\text{def}}{\Longrightarrow} N(Joy_a^{i=f(l,k)}(\gamma))$$

We consider that the agent happiness about a state of affairs has to be linearly proportional to its attitude about it. However, we want this intensity to evolve logarithmically according to the believability level of its occurrence. This way, we get to trigger more salient emotions even with relatively weak beliefs. Nevertheless, let us remind here that we only consider beliefs which levels are greater than a certain threshold we set as $mod\_thld = 0.5$ in our implementation. Then, we adjust the value in [0,1], calculate the intensity and readjust the result in [0.5,1] to get significant levels:

$$i = [k \times [(Ln((l - 0.5) \times 2) - Ln\_min)/ - Ln\_min]]/2 + 0.5$$

where $Ln\_min = Ln(x)$ when $x$ tends to 0, i.e. the smallest value of $Ln(x)$ coded by the machine.

All the other rules of Section 2.2.5 follow the same approach, generating emotions which intensities are linearly proportional to the agent's attitudes and logarithmically to its beliefs. Besides, regarding (2.11) and (2.24) the generated beliefs are linearly proportional to the initial beliefs about states of affairs and logarithmically to those about rules.

There are other level functions in our model that do not follow this approach, such as the one called in (2.28):

$$StrBel_a^l(\varphi) \wedge Att_a^k(F(\varphi)) \wedge Bel_a^{l'}(Att_b^{k'}(F(\varphi))) \wedge Like_{a,b}^h \wedge Dom_{a,b}^{h'} \stackrel{\text{def}}{\Longrightarrow} Att_a^{f(k,k',h,h')}(\varphi)$$

Here:

$$f(k, k', h, h') = k + \beta \frac{h - h'}{2} k'$$

This makes the agent's attitudes positively influenced by the interlocutors its like more than it dominates and vice versa. In our implementation, $\beta = k - k'$.

As for (2.37), we use the following:

$$f(k, k') = ((k + k')/4) + 0.5$$

## 3.4 Connection with the TARDIS project

An overview on TARDIS project's general architecture is presented in Figure 3.3. In this context, the module we implemented is part of the *Affective Module* that builds a model of beliefs and intentions about the user's mental states and about the course of actions in the ongoing interview [Anderson 13]. This component receives information about the user's mental states from the *Social cues interpretation Module* and provides the *Animation Module* with the agent's affects it has to express through both verbal and non-verbal behavior. It also communicates with the *Scenario Module* that controls the course of the interaction.

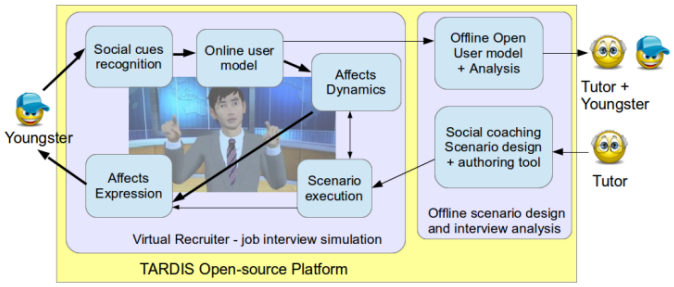



Figure 3.3: The TARDIS project general architecture [Anderson 13]

The communication between TARDIS components is managed by the C++ SEMAINE Application Programming Interface (API) that provides a multimodal dialogue system allowing for the implementation of social interaction capabilities such as emotional perception and non-verbal feedback. The Sensitive Artificial Listener paradigm on which this middleware is based makes real-time asynchronous communication possible. Therefore, each component must define a class that inherits from *semaine::components::Component* and then redefine at least one of the *Component.act()* and *Component.act(SEMAINEMessage *)* methods to respectively be able to send or receive messages. Beside, rich data is exchanged between components through eXtensible Markup Language (XML)-like files, based on representation like Functional Markup Language and Emotion Markup language. Please refer to the SEMAINE project web page[1] for more details.

Since the TARDIS project is compiled and executed on a Windows platform, we developped a Dynamic Link Library (DLL) that allows us to build an API that can be called by the Affective component to communicate the necessary inputs and outputs to the Prolog reasoner. This DLL has been created with the SWI-Prolog linker we presented in Section 3.2. Moreover, to allow the calling programs to run the reasoning engine, the prolog code has to be compiled in a stand-alone mode and called in the initialization. This provides a *program state* containing the initial database along with the prolog resources.

[1] http://www.semaine-project.eu/

## 3.5 Discussion

In this chapter, we presented the general architecture of our module allowing for theory of mind reasoning as well as details about its implementation and connection to the TARDIS project. Similar approach to connect this module to the MARC framework developed at LIMSI laboratory is intended. However, additional work is necessary in order to do that, as the latter project is implemented in Java language. Nevertheless, this would open the way to a large range of possible studies and applications for other contexts of interactions. Please refer to the MARC project web page[2] for more details.

[2]http://marc.limsi.fr/

# Chapter 4

# Evaluation

The last phase of this project aims at the evaluation of the model we designed and implemented. As we stated in the previous chapters, our main application context of interest is the job interviews simulation. In the following sections, we introduce the experimental protocol of study we conducted for this purpose and then discuss the results we obtained.

## 4.1 TARDISx: the job interview simulation

This experiment can be considered as a pre-test phase in the study of our work's possible contribution in the TARDIS project. This allows us to investigate the use of our theory of mind model in the context of job interviews with virtual recruiters.

### 4.1.1 Scenario

In this study, we made the subjects have a job interview with a virtual recruter. In this job interview, they would play the role of an unemployed youngster lacking work experience and applying for the job of sales department secretary. This candidate profile is one of the main targets of the TARDIS project. As for their personality, their education and professioanl background or the company in which they would be applying, the participants were free to imagine whatever they wanted to, as long as it was consistent with the scenario we proposed to them.

### 4.1.2 Method

We recruited 30 volunteers, 19 of them working or having an internship at LIMSI at that time, while the others were youngsters from outside the laboratory, a large majority of whom were graduate or PhD students in miscellaneous disciplines. So, all the subjects were aged over 24, had gone to university and were familiar with computers. 18 of them are native French speakers and the remaining have at least an intermediate French level.

The experiment consists in asking the participants to simulate a job interview with a virtual recruiter. This is done through a Graphical User Interface (GUI) that allows them to communicate with the agent (see Figure 4.1). In the beginning, a 2-question background survey is filled with information about subjects' knowledge and experience in job interviews. Then, the GUI is introduced to them along with the scenario detailed above. The whole preliminary part detailed above lasts from 5 to 10 minutes. At the end of the job interview, a 9-question evaluation questionnaire about interview difficulty, credibility and "pleasantness" is filled by the subjects.

There are 6 kinds of agents, each participant only interacts with one of them. We implemented 3 distinct recruiter profiles with our model: one that only asks regular questions, one that tries to make the candidates feel at ease and one that, contrariwise, asks embarassing questions. This is simply done by varying their goals regarding the emotional reaction they want to elicit. All of them have the same ToM reasoner described in Chapter 3, including the commonsense rules. To each profile, corresponds a placebo, i.e. an agent that apparently behaves in the same way but without any reasoning process. These agents just ask the same question as the corresponding profiles in a predefined and hard-coded way. The outputs, i.e. the affective state and the three evaluation bars, are just weighted sum of the inputs. The former just considers the valence of the candidate's affects and the latter weighs the input imitating some of the commonsense rules of the ToM agents but applying the same coefficients without regard to the topic it addressed.

In any case, even though the subjects are asked to answer the recruter by writing, the content is never taken into account. Although, when asking some questions considered as requiring elaborated answers in real job interviews, the ToM agents can refer to the awswers' length along with their evaluation of the candidate to decide whether to inquire more details.

#### Objectives

The main purpose of this experiment is to assess the impact of the ToM module on the quality of the interaction. Besides, it can allow us to compare the agent's profiles regarding their influence on the candidates' elicited emotional states and behavior.

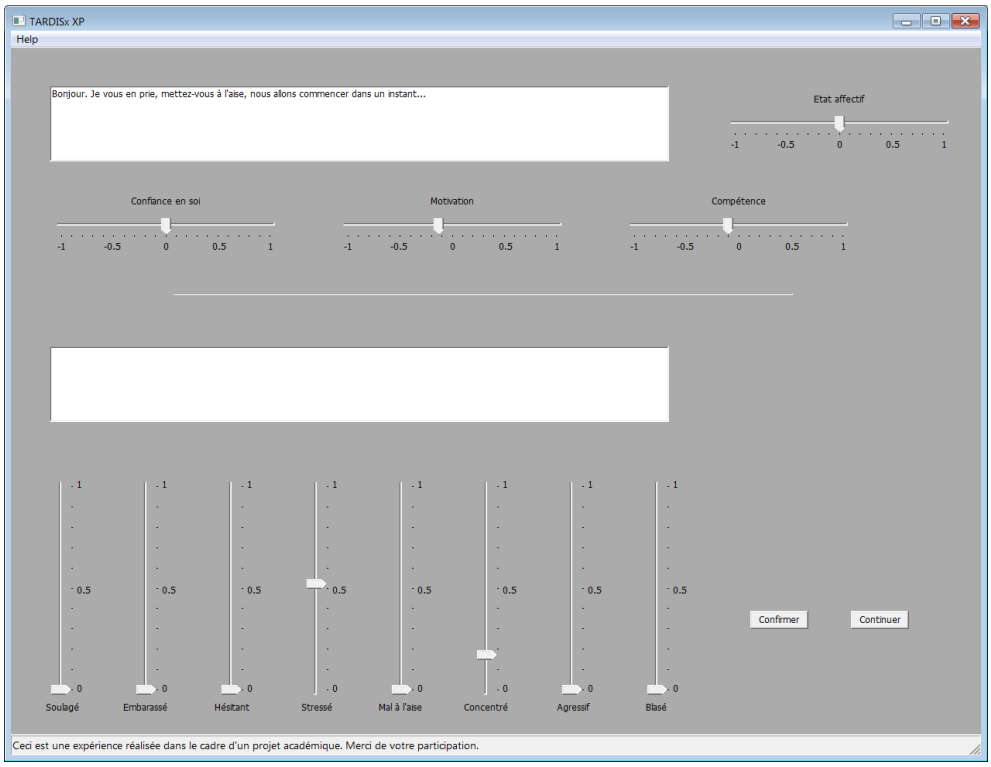



Figure 4.1: TARDISx experiment user interface

### Hypothesis

- H1: The interaction with ToM agents will seem more credible and agreeable than with placebos,
- H2: The profile variation will have an impact on the participants' emotional states,
- H3: The profile variation will have an impact on the participants' appreciation of the difficulty of the interview.

### Variables

- TOM: The recruiter has the ToM reasoner or is a placebo,
- PROFILE: The recruiter asks friendly (PROFILE_A), regular (PROFILE_B) or unpleasant (PROFILE_C) questions,
- XP: The candidate had less than five real job interviews before this study or more (five included),
- TRAINING: The candidate had some kind of training about how to behave in job interviews before this study or not.

### 4.1.3 Measures

**Evaluation questionnaire** (Subjective measures):

- DIFF: The candidate's evaluation of the interview's difficulty level,
- CRED_GLOB: The candidate's evaluation of the credibility of the recruiter's global behavior,
- CRED_AFF: The candidate's evaluation of the credibility of the recruiter's affective state variation
- CRED_CONF: The candidate's evaluation of the credibility of the recruiter's assessment of his/her self-confidence,
- CRED_MOTI: The candidate's evaluation of the credibility of the recruiter's assessment of his/her motivation,
- CRED_QUAL: The candidate's evaluation of the credibility of the recruiter's assessment of his/her qualitification,
- UNDERSTND: The candidate's impression on whether the recruiter took his/her written answers into account,
- EMPATHY: The candidate's impression on whether the recruiter took his/her emotional reactions into account,
- PLEASANT: The candidate's evaluation of the interaction's pleasantness.

**Interaction's log** (objective measures):

- TOT_TIME: The total duration of the interaction,
- AFF_TOT: The mean amount of information the candidate gave about his/her affective state,
- AFF_REL: The mean amount of information the candidate gave about his/her relief state,
- AFF_EMB: The mean amount of information the candidate gave about his/her embarrassment state,
- AFF_HES: The mean amount of information the candidate gave about his/her hesitation state,

- AFF_STR: The mean amount of information the candidate gave about his/her stress state,
- AFF_IAE: The mean amount of information the candidate gave about his/her uneasiness state,
- AFF_FOC: The mean amount of information the candidate gave about his/her focusing state,
- AFF_AGG: The mean amount of information the candidate gave about his/her aggressiveness state,
- AFF_BOR: The mean amount of information the candidate gave about his/her boredom state,

### 4.1.4 Results

In order to analyse the data collected in this study, we rely on statistical tests. There are two families of tests one can perform according to the nature of data: parametric and non-parametric. The former assume they follow parametric distributions, i.e. distributions that can be characterized by a set of parameters, and are more powerful on this sort of data. However, the latter allow for statistical analysis of any sort of data. Generally, normality tests are used to check the adequacy with the normal distribution – characterized by the mean and the variance – as it allows for modeling random natural phenomena. Shapiro–Wilks test shows that, except for TOT_TIME, none of objective and subjective measures follows a normal distribution. See Figure A.1. Therefore, in the following, we perform non-parametric analysis.

#### Relations between measures

To study the relations between our measures, we rely on the Spearman method that is used for non-parametric distributions. First, we test the bivariate correlation between the subjectives measure, i.e. the participants' answers. Thus, we find out that CRED_GLOB is highly correlated with UNDERSTND ($p < 0.05$) but with none of the other credibility measures (CRED_AFF, CRED_CONF, CRED_MOTI and CRED_QUAL), which suggests that the subjects assessment of the interaction's credibility is mainly based on the dialog quality. Indeed, a large majority of the subjects mentioned – often exclusively – the recruiter's questions (*"standard questions"*), their order, their redundancy, etc. in the explanatory field following this question. Besides, PLEASANT is significantly correlated with UNDERSTND as well as with EMPATHY ($p < 0.05$). On the other hand, there are several significant pairwise correlations among CRED_AFF, CRED_CONF, CRED_MOTI and CRED_QUAL. Cronbach's internal consistency test gives the coefficient $\alpha = 0.679$,

which although relatively low is acceptable given the number of items. This might allows us to consider that the theory of mind related credibility we are interested in in this study can be observed through an underlying factor carried by these four measures. Therefore, we define CRED_TOM as the mean value of CRED_AFF, CRED_CONF, CRED_MOTI and CRED_QUAL. This measure turns out to be correlated with PLEASANT ($p < 0.05$), which did not appear for the credibility measures taken separately, except for CRED_CONF ($p < 0.05$).

As for objective measures, there are significant pairwise correlations among AFF_EMB, AFF_HES, AFF_STR and AFF_IAE ($p < 0.05$) between AFF_STR and AFF_EMB and $p < 0.01$ otherwise). AFF_REL is also highly correlated to AFF_HES, AFF_STR and AFF_IAE ($p < 0.01$). Moreover, AFF_FOC and AFF_STR are significantly correlated ($p < 0.05$) as well as AFF_AGG and AFF_IAE ($p < 0.01$). Finally, AFF_BOR is correlated to AFF_AGG ($p < 0.01$), AFF_HES ($p < 0.01$) and AFF_EMB ($p < 0.05$).

Please refer to the correlations table shown in Figure A.4 for all the correlation coefficients values.

### Measures comparison

Since we have 30 independent samples sometimes divided in more than two groups, we use the Kruskal–Wallis method. This non-parametric method is used to see if the samples from the same group originate from the same distribution but cannot identify exactly where and how many differences occur. So, when significant results appear, the – non-parametric as well – Mann-Whitney test can be used to analyse the groups pairwise. See Figure A.2 for groups division.

No significant effect of TOM or XP on the measures appears. See Figure A.5. Which means that whether the agent's behavior is based on the ToM reasoner does not affect the participants affective states nor their evaluation of the interaction, neither the number of job interviews their had before this study. However, there is a main effect of TRAINING on AFF_FOC ($Chi^2(1, 629) = 6.340; p < 0.05$) and on AFF_EMB ($Chi^2(1, 629) = 4.181; p < 0.05$) with a tendency on CRED_CONF ($Chi^2(1, 269) = 3.640; p = 0.056$). Therefore, we perform the planned comparison Mann–Whitney test and see that subjects that were somehow trained for job interviews show more focused ($U = 46; p < 0.05$) and embarrassed ($U = 57; p < 0.05$) attitudes and tend to find the evaluation of their self-confidence more relevant ($U = 63.5; p = 0.056$) than those who were not.

Kruskal–Wallis also reveals a main effect of PROFILE on AFF_TOT ($Chi^2(2, 629) = 11.435; p < 0.01$) and particularly AFF_EMB ($Chi^2(2, 629) = 6.231; p < 0.05$) and AFF_FOC ($Chi^2(2, 629) = 9.218; p < 0.01$). Mann–Whitney then shows that participants that interact with PROFILE_A express more affects in general ($U = 20; p < 0.05$), more embarrassment ($U = 20; p < 0.05$) and more concentration ($U = 21; p < 0.05$)

than those who interact with PROFILE_B. Likewise, PROFILE_C elicits more affects ($U = 6; p < 0.01$) and in particular stress ($U = 18; p < 0.05$), uneasiness ($U = 24; p < 0.05$) and concentration ($U = 10; p < 0.01$) than PROFILE_B. We also note that in this case, no effect appears regarding embarrassment ($U = 26; p = 0.069$). Finally, no significant effect is revealed between PROFILE_A and PROFILE_C.

Please refer to A for more details about the results of Kruskal–Wallis and Mann–Whitney analysis.

## 4.2 Discussion

The correlations between the objective measures, i.e. the intensity of the participants' expressed attitudes, seem coherent and consistent with the context. For example, "embarrassed", "hesitant", "stressed" and "Ill at ease" are quite similar affects and it seems natural that their use was highly linked. Also, "stressed" and "focused" are the most obvious states one would be in in real job interviews, so they were often used together. Besides, as they are the most negative emotional feedback one can give in such a social context, the "bordom" and "aggressiveness" correlation makes sense. They were the least expressed affects and participants who exhibited them were probably either too honest about the unpleasantness they were experiencing or testing the system's limits. All in all, one can consider that participants were acting coherently rather than giving the recruiter random feedback.

Other results regarding the participants behavior seem interesting enough to be pointed out. For instance, neither the number of job interviews they had in the past nor the fact that they were trained influenced their behavior during the experiment or their post-hoc evaluation. This might suggest that even though the TARDIS project mainly targets unemployed and inexperienced youngsters, this tool could also benefit other kind of users. To this end, it should be flexible enough to adapt to their needs in terms of scenarios and recruiters' profiles variety and advisory feedback.

Additionally, our intuition suggested that the total interaction duration would be an immersion indicator and thus possibly correlated to the perceived credibility level. This is probably due to the fact that the kind of job the participants were asked to apply for did not match their career. The subjects' difficulty to adapt to the scenario is perhaps the main factor that influenced the total duration. One should also note, as shown in Figure A.3 that the use of ToM reasoning and the kind of profile did not have any significant impact on this measure either, although these variables have a direct effect on the number of iterations in the interaction. This consequently seems consistent with the former assumption.

Regarding the hypothesis we formulated in the previous section, H3 did not verify. Indeed, the recruiter profile did not have any effect on the participants' appreciation of the

difficulty. Although they were designed with the idea that the less unpleasant questions were asked the easier the interview would be, it was not perceived this way by the subjects. This might be explained by the agent's affective state. Indeed, if a PROFILE_C agent asks a destabilizing question and elicits hesitation or embarrassment, it achieves one of its subgoals which triggers a positive emotion. This variation can be perceived by the participants as an attempt to appear friendlier and reassuring, thus removing the impression of difficulty. This raises the question of whether the virtual recruiter should hide its "real" emotional state and of the kind of feedback would the candidates have on their performance then.

Nevertheless, as far as H2 is concerned, the profiles appeared to have an impact on the elicited affects' intensity. The comparative analysis showed that asking regular questions elicited less emotional reactions in the subjects. On the other hand, no matter the valence, the more the virtual recruiter tried to elicit emotional reactions, the more it succeeded. This is an interesting result for the TARDIS project. It confirms that some questions should have a direct impact on the amount of social signals that would be expressed and be potentially detectable by the system. Also, it confirms that several profiles and behaviors should be implemented to test it.

Finally, our first hypothesis, and a priori most relevant regarding the evaluation of our module, have not been verified. Whether the agent's behavior was based on a theory of mind reasoning did not influence the impression of credibility or pleasantness, nor any of our measures in general. There are several ways of explaining this result. First of all, although no reasoning process is used in the placebos, the latter do imitate the corresponding ToM agents. The differences between their behaviors are quite subtle. Since each subject only interacted with one recruiter, placebos might, for instance, have been perceived as empathic but still credible recruters, or, at least, not less credible than ToM agents. Besides, correlations between subjective measures revealed that the perceived credibility was mostly related to the feeling that the recruiter was taking the written answers into account. Hence, in the evaluation phase, the notion of credibility remained quite superficial. The one we intended to elicit – i.e. a ToM-related credibility relying on emotional reactivity and on cognitive assessment of the other's behavior – and that we represented by the mean value CRED_TOM was less salient. This one was related to the interaction's pleasantness.

Yet, in a certain way, the results about the impact of profile variation on participants' emotional reactions can be considered as significant for the assessment of the theory of mind module as well. Indeed, the selection of some questions relies on the reasoning about the mental and emotional states they could induce, regardless of whether it is performed online by the agent or hard-coded in order to imitiate the online reasoning process. So, the more the recruter asks such questions, the more mindreading it performs. Consequently, this could mean there is a relation between the use of ToM and the intensity of elicited affective attitudes.

In conclusion, we believe that the lack of significant results regarding the influence of our module is due to a few shortcomings that the experimental protocol suffers. For

instance, not to mention the small number of participants as compared to the number of experimental conditions, the GUI is not very user-friendly and was not expected to make the interactions seem remarkably pleasant. One could assume that with a 3D virtual agent, voice recognition and non-verbal communication, the influence of our module would have been more salient. Hence, the need to evaluate our work in a TARDIS prototype testing and, if possible, in others studies based on MARC virtual agents. Moreover, theory of mind is a complex process the relies on various other cognitive and perceptual processes. Since we do not know the exact underlying functionning in human being, it is not only hard to model but also to assess. In the litterature, there are validated methods to evaluate whether subjects – generally children – have it and use it. [Blijd-Hoogewys 08], for instance, presents a set of storybooks that allows for the study of ToM development through related tasks. Nevertheless, there is no interactional aspect in these tests and we cannot base the evaluation of our model on them. From the computational point of view, [Harbers 11] points out the issue of evaluating a ToM model. In this work, the course of events and the agent's actions and explanations are specified in advance for different scenarios. Thus, the ST and TT ToM models are evaluated based on whether they match these specifications. Similarly, [Pynadath 13] builds expectations about user's actions – based on formal models in the specific context of wartime negotiations – in order to model a simplified theory of mind and then compare them with the actual user's behavior. These two approaches are not applicable in our case, as it is much more complicated to construct such specifications in the context of interactions like those we are interested in. However, the design of a psychologically validated protocol that could evaluate our model is under discussion.

# Conclusion

The purpose of this work is to investigate the influence and the contribution of an *emotion*-oriented *Theory of Mind* module in Human/Agent *interactions*. Theory of Mind or *mindreading* are the terms used when it comes to the ability of human – and eventually non-human – beings to interpret, explain or predict others' behavior. This social phenomenon has been widely examined by philosophers [Botterill 99] [Goldman 06], psychologists [Wimmer 83] [Leslie 94] [Baron-Cohen 97], neuroscientists [Vogeley 01], etc. Addressing this topic in the context of *Affective Computing* aims at the study of the theories presented in such disciplines from a more practical point of view – through computational implementation – as well as the development of computer systems that would be able to interact with humans in a more fluent and efficient way [Scassellati 02] [Peters 05] [Pynadath 05] [Aylett 08] [Harbers 11] [Bosse 11].

From the theoretical point of view, theorists and simulationists debated for a long time about whether the theory of mind was based on a set of rules one learns about the functioning of human mind or on a projection process that lets us take others' perspective. But, the hybrid approach we adopted argues in favor of a combination of both mechanisms [Botterill 99] [Goldman 06] [Vogeley 01]. Nevertheless, as we work on symbolic Artificial Intelligence, we did give more importance to folk-psychology and commonsense reasoning. Indeed, the project's first phase fathered a logical framework that allows for modeling human and virtual agents' mental states – through a Beliefs Desires and Intentions (BDI)-based approach [Bratman 99] [Rao 91] –, communication – based on speech acts theory [Austin 62] [Searle 69] –, and social relations and interactions [Leary 57] [Kiesler 96] [Castelfranchi 97]. As for the affective aspect, we relied on appraisal theories of emotions that defend a cognitive evaluation of states of affairs [Scherer 10] [Ortony 90]. The resulting non-domain-specific formal model is potentially adaptable to any context of interaction. The implementation of its logic has been done in the Prolog declarative programming language.

Additionally, this work is part of the TARDIS projet that aims to develop an open-source online and offline social training platform and to facilitate youngsters' – mainly those at risk of social exclusion – access to employment [Anderson 13]. The integration of our model to the TARDIS Affective Module would allow the agent to reason about the

human's mental and emotional states and to influence the course of interaction.

Because of the connection to the TARDIS projet, job interviews simulation is our main application field of interest. Once the reasoning engine implemented for this purpose, the third phase of our work consisted in the evaluation of the stand-alone version. This preliminary experiment gave interesting results, for example regarding the influence of the recruiter's profile variation on the elicited emotional attitudes in the human candidate. On the other hand, our most important hypothesis – stating that interaction relying on the theory of mind module would be perceived as more credible and pleasant – did not verify. Yet, we believe that these results are mainly the consequence of some weaknesses in the experimental protocol such as the insufficient number of participants, the lack of user-friendliness in the Graphical User Interface and the design of the placebo agents. In any case, our agents do present a clear advantage compared to non-reasoning ones, which is the explanatory aspect. Indeed, at the end of an interaction, the user can have access to the agent's mental states, reasoning and behavior history and use it to understand what made it act like it did. This is a very useful feature in the context of youngsters training for job interviews for example.

Unfortunately, due to time contraints, we have not been able to test our module functioning as part of a TARDIS prototype yet. This is an ongoing task that should benefit from the preliminary study's results as well as compensate for some of its shortcomings. A connection with the MARC virtual agents could also open the way for a lot of interesting studies regarding Human/Agent interactions. As for the evaluation of the theory of mind model, the design of a recognition test protocol similar to [Blijd-Hoogewys 08], that would be validated from a psychological perspective, is also under discussion.

# Bibliography


[Adam 09] Carole Adam, Andreas Herzig & Dominique Longin. *A logical formalization of the OCC theory of emotions*. Synthese, vol. 168, no. 2, pages 201–248, February 2009.

[Adam 12] Carole Adam & Dominique Longin. *Honte ou culpabilité? (That is the question)*. In Proceedings of the workshop Affect, Compagnon Artificiel, Interaction (WACAI), 2012.

[Anderson 13] K. Anderson, E. André, T. Baur, S. Bernardini, M. Chollet, E. Chryssafidou, I. Damian, C. Ennis, A. Egges, P. Gebhard, H. Jones, M. Ochs, C. Pelachaud, K. Porayska-Pomsta, P. Rizzo & N. Sabouret. *The TARDIS framework: intelligent virtual agents for social coaching in job interviews*. In Proceedings of the 10th international conference on Advances in Computer Entertainment technology, 2013. To be published.

[Arnold 60] Magda B Arnold. Emotion and Personality: Psychological aspects, volume 1. Columbia University Press, 1960.

[Austin 62] John L Austin. How To Do Things With Words. Oxford Paperbacks Philosophy. Oxford University Press, 1962.

[Aylett 08] Ruth Aylett & Sandy Louchart. *If I were you: double appraisal in affective agents*. In Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3, pages 1233–1236. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

[Baron-Cohen 97] Simon Baron-Cohen. Mindblindness: An essay on autism and theory of mind. MIT press, 1997.

[Blijd-Hoogewys 08] E M A Blijd-Hoogewys, P L C Van Geert, M Serra & R B Minderaa. *Measuring theory of mind in children. Psychometric properties of the ToM storybooks*. Journal of Autism and Developmental Disorders, vol. 38, no. 10, pages 1907–1930, 2008.

[Bosse 11] Tibor Bosse, Zulfiqar A Memon & Jan Treur. *A recursive BDI agent model for Theory of Mind and its applications.* Applied Artificial Intelligence, vol. 25, no. 1, pages 1–44, 2011.

[Botterill 99] George Botterill & Peter Carruthers. The philosophy of psychology. Cambridge University Press, 1999.

[Bratman 99] Michael E Bratman. Intention, plans, and practical reason. Cambridge University Press, 1999.

[Castelfranchi 97] Cristiano Castelfranchi. *Modelling social action for AI agents.* IJCAI'97 Proceedings of the Fifteenth international joint conference on Artifical intelligence - Volume 2, vol. 103, no. 1, pages 1567–1576, 1997.

[Castelfranchi 98] Cristiano Castelfranchi & Rino Falcone. *Towards a theory of delegation for agent-based systems.* Robotics and Autonomous Systems, vol. 24, no. 3, pages 141–157, 1998.

[Dastani 12] Mehdi Dastani & Emiliano Lorini. *A logic of emotions : from appraisal to coping.* In Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 2, pages 1133–1140, 2012.

[Davis 79] Steven Davis. *Perlocutions.* Linguistics and Philosophy, vol. 3, no. 2, pages 225–243, 1979.

[FIPA 02] Foundation for Intelligent Physical Agents FIPA. *Communicative act library specification.* http://www.fipa.org – Accessed:09/08/2013, 2002.

[Goldman 06] Alvin I Goldman. Simulating minds: The philosophy, psychology, and neuroscience of mindreading. Oxford University Press, 2006.

[Guiraud 11] Nadine Guiraud, Dominique Longin, Emiliano Lorini, Sylvie Pesty & Jérémy Rivière. *The face of emotions : a logical formalization of expressive speech acts.* In The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 3, pages 1031–1038, 2011.

[Harbers 11] Maaike Harbers. *Explaining agent behavior in virtual training.* SIKS dissertation series, vol. 2011, no. 35, 2011.

[Herzig 02] Andreas Herzig & Dominique Longin. *A logic of intention with cooperation principles and with assertive speech acts as communication primitives.* In Proceedings of the first international joint conference on

Autonomous agents and multiagent systems: part 2, pages 920–927. ACM, 2002.

[Kiesler 96] Donald J Kiesler. Contemporary interpersonal theory and research: Personality, psychopathology, and psychotherapy. John Wiley & Sons, 1996.

[Lazarus 91] Richard S Lazarus. Emotion and adaptation. Oxford University Press New York, 1991.

[Leary 57] Timothy Francis Leary*et al.* Interpersonal diagnosis of personality. Ronald Press New York, 1957.

[Leslie 87] Alan M Leslie. *Pretense and representation: The origins of "Theory of Mind".* Psychological review, vol. 94, no. 4, pages 412–426, 1987.

[Leslie 94] Alan M Leslie. *ToMM, ToBy, and Agency: Core architecture and domain specificity.* Mapping the mind: Domain specificity in cognition and culture, pages 119–148, 1994.

[Malle 99] Bertram F Malle. *How people explain behavior: A new theoretical framework.* Personality and Social Psychology Review, vol. 3, no. 1, pages 23–48, 1999.

[Marcu 00] Daniel Marcu. *Perlocutions: The Achilles' heel of speech act theory.* Journal of pragmatics, vol. 32, no. 12, pages 1719–1741, 2000.

[Nichols 03] Shaun Nichols & Stephen P Stich. Mindreading: An integrated account of pretence, self-awareness, and understanding other minds. Oxford University Press Oxford, 2003.

[Ochs 09] Magalie Ochs, Nicolas Sabouret & Vincent Corruble. *Simulation of the Dynamics of Nonplayer Characters' Emotions and Social Relations in Games.* Computational Intelligence and AI in Games, IEEE Transactions on, vol. 1, no. 4, pages 281–297, 2009.

[Ortony 90] Andrew Ortony, Gerald L Clore & Allan Collins. *The Cognitive Structure of Emotions*, 1990.

[Peters 05] Christopher Peters. *Foundations of an Agent Theory of Mind Model for Conversation Initiation in Virtual Environments.* Virtual Social Agents, page 163, 2005.

[Pynadath 02] David V Pynadath & Milind Tambe. *Multiagent teamwork: Analyzing the optimality and complexity of key theories and models.* In Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2, pages 873–880. ACM, 2002.

[Pynadath 05] David V Pynadath & Stacy C Marsella. *PsychSim: modeling theory of mind with decision-theoretic agents*. In International Joint Conference on Artificial Intelligence, volume 19, pages 1181–1186, 2005.

[Pynadath 13] David V Pynadath, Ning Wang & Stacy C Marsella. *Are you thinking what I'm thinking? An Evaluation of a Simplified Theory of Mind*. In Intelligent Virtual Agents, pages 44–57. Springer, 2013.

[Rao 91] Anand S Rao & Michael P Georgeff. *Modeling Rational Agents within a BDI-Architecture*. In Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning, 1991.

[Rao 95] Anand S Rao, Michael P Georgeff & Others. *BDI agents: From theory to practice*. In Proceedings of the first international conference on multi-agent systems (ICMAS-95), pages 312–319. San Francisco, 1995.

[Scassellati 02] Brian Scassellati. *Theory of Mind for a Humanoid Robot*. Autonomous Robots, vol. 12, pages 13–24, 2002.

[Scherer 01] Klaus R Scherer. *Appraisal considered as a process of multilevel sequential checking*. Appraisal processes in emotion: Theory, methods, research, pages 92–120, 2001.

[Scherer 10] Klaus R Scherer. *Emotion and emotional competence: conceptual and theoretical issues for modelling agents*. Blueprint for Affective Computing, pages 3–20, 2010.

[Searle 69] John R Searle. Speech acts: An essay in the philosophy of language, volume 626. Cambridge university press, 1969.

[Searle 76] John R Searle. *A classification of illocutionary acts*. Language in society, vol. 5, no. 01, pages 1–23, 1976.

[Vogeley 01] Kai Vogeley, P Bussfeld, Albert Newen, S Herrmann, F Happe, P Falkai, W Maier, N J Shah, G R Fink & K Zilles. *Mind reading: neural mechanisms of theory of mind and self-perspective*. Neuroimage, vol. 14, no. 1, pages 170–181, 2001.

[Wellman 90] Henry M Wellman & Jacqueline D Woolley. *From simple desires to ordinary beliefs: The early development of everyday psychology*. Cognition, vol. 35, no. 3, pages 245–275, 1990.

[Wimmer 83] Heinz Wimmer & Josef Perner. *Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception*. Cognition, vol. 13, no. 1, pages 103–128, 1983.

# Appendices

# Appendix A

# TARDISx analysis results

This appendix presents the detailed results of the TARDISx experiment's analysis.

**Tests of Normality**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| TOT_TIME | ,131 | 30 | ,200* | ,971 | 30 | ,578 |
| CRED_GLB | ,272 | 30 | ,000 | ,845 | 30 | ,000 |
| CRED_AFF | ,201 | 30 | ,003 | ,904 | 30 | ,011 |
| CRED_CONF | ,290 | 30 | ,000 | ,825 | 30 | ,000 |
| CRED_MOTI | ,246 | 30 | ,000 | ,904 | 30 | ,010 |
| CRED_QUAL | ,250 | 30 | ,000 | ,884 | 30 | ,003 |
| CRED_TOM | ,172 | 30 | ,024 | ,951 | 30 | ,181 |
| DIFF | ,259 | 30 | ,000 | ,888 | 30 | ,004 |
| UNDERSTND | ,204 | 30 | ,003 | ,878 | 30 | ,003 |
| EMPATHY | ,316 | 30 | ,000 | ,840 | 30 | ,000 |
| PLEASANT | ,202 | 30 | ,003 | ,908 | 30 | ,013 |
| AFF_TOT | ,191 | 30 | ,007 | ,884 | 30 | ,003 |
| AFF_Rel | ,137 | 30 | ,155 | ,925 | 30 | ,037 |
| AFF_Emb | ,262 | 30 | ,000 | ,823 | 30 | ,000 |
| AFF_Hes | ,212 | 30 | ,001 | ,775 | 30 | ,000 |
| AFF_Str | ,237 | 30 | ,000 | ,815 | 30 | ,000 |
| AFF_Unc | ,257 | 30 | ,000 | ,775 | 30 | ,000 |
| AFF_Foc | ,114 | 30 | ,200* | ,943 | 30 | ,111 |
| AFF_Agg | ,336 | 30 | ,000 | ,471 | 30 | ,000 |
| AFF_Bor | ,283 | 30 | ,000 | ,635 | 30 | ,000 |

a. Lilliefors Significance Correction
*. This is a lower bound of the true significance.

Figure A.1: Shapiro-Wilk normality test results for all the measures

**Between-Subjects Factors**

| | | N |
|---|---|---|
| TRAINING | 1 | 11 |
| | 2 | 19 |
| XP | 1 | 14 |
| | 2 | 16 |
| TOM | 1 | 15 |
| | 2 | 15 |
| PROFILE | 1 | 10 |
| | 2 | 10 |
| | 3 | 10 |

Figure A.2: Inter-subject factors

**Tests of Between-Subjects Effects**

Dependent Variable:TOT_TIME

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Corrected Model | 6042634,500 | 17 | 355449,088 | 1,545 | ,224 |
| Intercept | 52837089,03 | 1 | 52837089,03 | 229,676 | ,000 |
| TRAINING | 55211,650 | 1 | 55211,650 | ,240 | ,633 |
| XP | 12084,250 | 1 | 12084,250 | ,053 | ,823 |
| TOM | 428615,637 | 1 | 428615,637 | 1,863 | ,197 |
| PROFILE | 1487643,850 | 2 | 743821,925 | 3,233 | ,075 |
| TRAINING * XP | 11890,286 | 1 | 11890,286 | ,052 | ,824 |
| TRAINING * TOM | 484410,083 | 1 | 484410,083 | 2,106 | ,172 |
| TRAINING * PROFILE | 1068045,843 | 2 | 534022,921 | 2,321 | ,141 |
| XP * TOM | 87089,338 | 1 | 87089,338 | ,379 | ,550 |
| XP * PROFILE | 185646,704 | 2 | 92823,352 | ,403 | ,677 |
| TOM * PROFILE | 28241,800 | 2 | 14120,900 | ,061 | ,941 |
| TRAINING * XP * TOM | ,000 | 0 | . | . | . |
| TRAINING * XP * PROFILE | ,000 | 0 | . | . | . |
| TRAINING * TOM * PROFILE | ,000 | 0 | . | . | . |
| XP * TOM * PROFILE | 402637,103 | 1 | 402637,103 | 1,750 | ,211 |
| TRAINING * XP * TOM * PROFILE | ,000 | 0 | . | . | . |
| Error | 2760603,667 | 12 | 230050,306 | | |
| Total | 99127339,00 | 30 | | | |
| Corrected Total | 8803238,167 | 29 | | | |

a. R Squared = ,686 (Adjusted R Squared = ,242)

Figure A.3: ANOVA analysis for the variable's effects on the total interaction duration

**Correlations**

| | | | CRED_GLB | CRED_AFF | CRED_CONF | CRED_MOTI | CRED_QUAL | CRED_TOM | DIFF | UNDERSTND | EMPATHY | PLEASANT | AFF_TOT | AFF_Rel | AFF_Emb | AFF_Hes | AFF_Str | AFF_Unc | AFF_Foc | AFF_Agg | AFF_Bor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Spearman's rho | CRED_GLB | Correlation Coefficient | 1,000 | ,006 | ,075 | ,147 | ,097 | ,027 | ,144 | ,448* | -,049 | ,203 | -,274 | ,086 | ,021 | ,084 | -,111 | -,229 | -,409* | -,103 | -,203 |
| | | Sig. (2-tailed) | . | ,973 | ,693 | ,437 | ,609 | ,888 | ,448 | ,013 | ,799 | ,283 | ,143 | ,651 | ,914 | ,658 | ,561 | ,223 | ,025 | ,587 | ,282 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | CRED_AFF | Correlation Coefficient | ,006 | 1,000 | ,308 | -,007 | ,642** | ,658** | ,173 | ,234 | ,161 | ,310 | ,169 | ,029 | -,040 | -,026 | ,116 | ,040 | ,271 | -,165 | -,229 |
| | | Sig. (2-tailed) | ,973 | . | ,097 | ,971 | ,000 | ,000 | ,362 | ,213 | ,397 | ,095 | ,372 | ,880 | ,833 | ,893 | ,543 | ,833 | ,147 | ,383 | ,224 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | CRED_CONF | Correlation Coefficient | ,075 | ,308 | 1,000 | ,208 | ,179 | ,492** | ,061 | ,250 | ,243 | ,401* | -,071 | -,187 | ,197 | -,277 | -,324 | -,002 | ,060 | -,004 | -,321 |
| | | Sig. (2-tailed) | ,693 | ,097 | . | ,270 | ,343 | ,006 | ,748 | ,183 | ,195 | ,028 | ,708 | ,322 | ,297 | ,138 | ,081 | ,991 | ,754 | ,981 | ,083 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | CRED_MOTI | Correlation Coefficient | ,147 | -,007 | ,208 | 1,000 | ,361* | ,561** | -,216 | ,214 | ,246 | ,309 | -,303 | -,174 | -,223 | -,520** | -,499** | -,354 | -,162 | ,023 | -,281 |
| | | Sig. (2-tailed) | ,437 | ,971 | ,270 | . | ,050 | ,001 | ,251 | ,256 | ,190 | ,097 | ,103 | ,358 | ,236 | ,003 | ,005 | ,055 | ,393 | ,903 | ,132 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | CRED_QUAL | Correlation Coefficient | ,097 | ,642** | ,179 | ,361* | 1,000 | ,866** | ,120 | ,208 | ,079 | ,299 | ,256 | ,014 | -,149 | ,002 | ,026 | -,101 | ,238 | ,063 | -,156 |
| | | Sig. (2-tailed) | ,609 | ,000 | ,343 | ,050 | . | ,000 | ,528 | ,270 | ,678 | ,108 | ,173 | ,940 | ,433 | ,991 | ,893 | ,596 | ,205 | ,739 | ,410 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | CRED_TOM | Correlation Coefficient | ,027 | ,658** | ,492** | ,561** | ,866** | 1,000 | ,068 | ,225 | ,253 | ,450* | ,031 | -,128 | -,183 | -,234 | -,206 | -,180 | ,141 | -,047 | -,352 |
| | | Sig. (2-tailed) | ,888 | ,000 | ,006 | ,001 | ,000 | . | ,723 | ,231 | ,177 | ,013 | ,869 | ,500 | ,334 | ,212 | ,274 | ,341 | ,458 | ,807 | ,057 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | DIFF | Correlation Coefficient | ,144 | ,173 | ,061 | -,216 | ,120 | ,068 | 1,000 | -,116 | -,212 | -,243 | ,189 | ,287 | ,140 | ,389* | ,120 | ,366* | ,036 | ,090 | ,000 |
| | | Sig. (2-tailed) | ,448 | ,362 | ,748 | ,251 | ,528 | ,723 | . | ,543 | ,261 | ,196 | ,318 | ,124 | ,460 | ,034 | ,526 | ,047 | ,852 | ,638 | ,998 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | UNDERSTND | Correlation Coefficient | ,448* | ,234 | ,250 | ,214 | ,208 | ,225 | -,116 | 1,000 | ,163 | ,431* | -,125 | -,104 | -,012 | -,020 | ,008 | -,099 | -,197 | -,412* | -,462* |
| | | Sig. (2-tailed) | ,013 | ,213 | ,183 | ,256 | ,270 | ,231 | ,543 | . | ,389 | ,017 | ,510 | ,586 | ,948 | ,915 | ,966 | ,604 | ,297 | ,024 | ,010 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | EMPATHY | Correlation Coefficient | -,049 | ,161 | ,243 | ,246 | ,079 | ,253 | -,212 | ,163 | 1,000 | ,418* | ,081 | -,126 | -,214 | -,372* | -,156 | -,154 | ,294 | -,266 | -,414* |
| | | Sig. (2-tailed) | ,799 | ,397 | ,195 | ,190 | ,678 | ,177 | ,261 | ,389 | . | ,021 | ,672 | ,505 | ,256 | ,043 | ,410 | ,416 | ,115 | ,155 | ,023 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | PLEASANT | Correlation Coefficient | ,203 | ,310 | ,401* | ,309 | ,299 | ,450* | -,243 | ,431* | ,418* | 1,000 | ,024 | -,404* | -,072 | -,238 | -,078 | -,203 | ,143 | -,286 | -,523** |
| | | Sig. (2-tailed) | ,283 | ,095 | ,028 | ,097 | ,108 | ,013 | ,196 | ,017 | ,021 | . | ,899 | ,027 | ,705 | ,205 | ,680 | ,282 | ,452 | ,126 | ,003 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | AFF_TOT | Correlation Coefficient | -,274 | ,169 | -,071 | -,303 | ,256 | ,031 | ,189 | -,125 | ,081 | ,024 | 1,000 | ,371* | ,372* | ,495** | ,693** | ,591** | ,818** | ,360 | ,250 |
| | | Sig. (2-tailed) | ,143 | ,372 | ,708 | ,103 | ,173 | ,869 | ,318 | ,510 | ,672 | ,899 | . | ,043 | ,043 | ,005 | ,000 | ,001 | ,000 | ,051 | ,182 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | AFF_Rel | Correlation Coefficient | ,086 | ,029 | -,187 | -,174 | ,014 | -,128 | ,287 | -,104 | -,126 | -,404* | ,371* | 1,000 | ,190 | ,557** | ,453* | ,450* | ,027 | ,352 | ,406* |
| | | Sig. (2-tailed) | ,651 | ,880 | ,322 | ,358 | ,940 | ,500 | ,124 | ,586 | ,505 | ,027 | ,043 | . | ,314 | ,001 | ,012 | ,013 | ,888 | ,057 | ,026 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | AFF_Emb | Correlation Coefficient | ,021 | -,040 | ,197 | -,223 | -,149 | -,183 | ,140 | -,012 | -,214 | -,072 | ,372* | ,190 | 1,000 | ,392* | ,336 | ,484** | ,139 | ,364* | ,401* |
| | | Sig. (2-tailed) | ,914 | ,833 | ,297 | ,236 | ,433 | ,334 | ,460 | ,948 | ,256 | ,705 | ,043 | ,314 | . | ,032 | ,070 | ,007 | ,464 | ,048 | ,028 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | AFF_Hes | Correlation Coefficient | ,084 | -,026 | -,277 | -,520** | ,002 | -,234 | ,389* | -,020 | -,372* | -,238 | ,495** | ,557** | ,392* | 1,000 | ,540** | ,483** | ,071 | ,221 | ,477** |
| | | Sig. (2-tailed) | ,658 | ,893 | ,138 | ,003 | ,991 | ,212 | ,034 | ,915 | ,043 | ,205 | ,005 | ,001 | ,032 | . | ,002 | ,007 | ,710 | ,240 | ,008 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | AFF_Str | Correlation Coefficient | -,111 | ,116 | -,324 | -,499** | ,026 | -,206 | ,120 | ,008 | -,156 | -,078 | ,693** | ,453* | ,336 | ,540** | 1,000 | ,658** | ,391* | ,287 | ,239 |
| | | Sig. (2-tailed) | ,561 | ,543 | ,081 | ,005 | ,893 | ,274 | ,526 | ,966 | ,410 | ,680 | ,000 | ,012 | ,070 | ,002 | . | ,000 | ,032 | ,123 | ,203 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | AFF_Unc | Correlation Coefficient | -,229 | ,040 | -,002 | -,354 | -,101 | -,180 | ,366* | -,099 | -,154 | -,203 | ,591** | ,450* | ,484** | ,483** | ,658** | 1,000 | ,257 | ,536** | ,232 |
| | | Sig. (2-tailed) | ,223 | ,833 | ,991 | ,055 | ,596 | ,341 | ,047 | ,604 | ,416 | ,282 | ,001 | ,013 | ,007 | ,007 | ,000 | . | ,170 | ,002 | ,218 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | AFF_Foc | Correlation Coefficient | -,409* | ,271 | ,060 | -,162 | ,238 | ,141 | ,036 | -,197 | ,294 | ,143 | ,818** | ,027 | ,139 | ,071 | ,391* | ,257 | 1,000 | ,198 | ,078 |
| | | Sig. (2-tailed) | ,025 | ,147 | ,754 | ,393 | ,205 | ,458 | ,852 | ,297 | ,115 | ,452 | ,000 | ,888 | ,464 | ,710 | ,032 | ,170 | . | ,294 | ,682 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | AFF_Agg | Correlation Coefficient | -,103 | -,165 | -,004 | ,023 | ,063 | -,047 | ,090 | -,412* | -,266 | -,286 | ,360 | ,352 | ,364* | ,221 | ,287 | ,536** | ,198 | 1,000 | ,546** |
| | | Sig. (2-tailed) | ,587 | ,383 | ,981 | ,903 | ,739 | ,807 | ,638 | ,024 | ,155 | ,126 | ,051 | ,057 | ,048 | ,240 | ,123 | ,002 | ,294 | . | ,002 |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | AFF_Bor | Correlation Coefficient | -,203 | -,229 | -,321 | -,281 | -,156 | -,352 | ,000 | -,462* | -,414* | -,523** | ,250 | ,406* | ,401* | ,477** | ,239 | ,232 | ,078 | ,546** | 1,000 |
| | | Sig. (2-tailed) | ,282 | ,224 | ,083 | ,132 | ,410 | ,057 | ,998 | ,010 | ,023 | ,003 | ,182 | ,026 | ,028 | ,008 | ,203 | ,218 | ,682 | ,002 | . |
| | | N | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Figure A.4: Correlations table for all the measures

**Test Statistics[a,b]**

| | CRED_GLB | CRED_AFF | CRED_CONF | CRED_MOTI | CRED_QUAL | CRED_TOM | DIFF | UNDERSTND | EMPATHY | PLEASANT | AFF_TOT | AFF_Rel | AFF_Emb | AFF_Hes | AFF_Str | AFF_Unc | AFF_Foc | AFF_Agg | AFF_Bor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chi-Square | ,801 | ,250 | 3,640 | ,008 | ,290 | ,518 | ,041 | ,129 | 2,540 | ,458 | 2,745 | 1,401 | 4,184 | ,419 | ,104 | ,048 | 6,340 | ,044 | ,997 |
| df | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Asymp. Sig. | ,371 | ,617 | ,056 | ,927 | ,590 | ,472 | ,840 | ,720 | ,111 | ,498 | ,098 | ,237 | ,041 | ,518 | ,747 | ,826 | ,012 | ,833 | ,318 |

a. Kruskal Wallis Test
b. Grouping Variable: TRAINING

**Test Statistics[a,b]**

| | CRED_GLB | CRED_AFF | CRED_CONF | CRED_MOTI | CRED_QUAL | CRED_TOM | DIFF | UNDERSTND | EMPATHY | PLEASANT | AFF_TOT | AFF_Rel | AFF_Emb | AFF_Hes | AFF_Str | AFF_Unc | AFF_Foc | AFF_Agg | AFF_Bor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chi-Square | ,466 | ,213 | 1,164 | 1,876 | ,642 | ,300 | 1,175 | ,000 | 1,295 | ,171 | 2,117 | 2,240 | ,443 | ,021 | ,837 | ,016 | 2,906 | ,774 | 1,202 |
| df | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Asymp. Sig. | ,495 | ,645 | ,281 | ,171 | ,423 | ,584 | ,278 | ,983 | ,255 | ,679 | ,146 | ,134 | ,506 | ,884 | ,360 | ,899 | ,088 | ,379 | ,273 |

a. Kruskal Wallis Test
b. Grouping Variable: XP

**Test Statistics[a,b]**

| | CRED_GLB | CRED_AFF | CRED_CONF | CRED_MOTI | CRED_QUAL | CRED_TOM | DIFF | UNDERSTND | EMPATHY | PLEASANT | AFF_TOT | AFF_Rel | AFF_Emb | AFF_Hes | AFF_Str | AFF_Unc | AFF_Foc | AFF_Agg | AFF_Bor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chi-Square | ,101 | ,002 | ,061 | 1,214 | ,269 | ,541 | ,135 | ,030 | ,143 | ,613 | ,011 | ,759 | ,951 | ,191 | ,269 | ,945 | ,073 | 1,647 | 1,196 |
| df | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Asymp. Sig. | ,751 | ,965 | ,805 | ,270 | ,604 | ,462 | ,713 | ,863 | ,705 | ,434 | ,917 | ,384 | ,329 | ,662 | ,604 | ,331 | ,787 | ,199 | ,274 |

a. Kruskal Wallis Test
b. Grouping Variable: TOM

**Test Statistics[a,b]**

| | CRED_GLB | CRED_AFF | CRED_CONF | CRED_MOTI | CRED_QUAL | CRED_TOM | DIFF | UNDERSTND | EMPATHY | PLEASANT | AFF_TOT | AFF_Rel | AFF_Emb | AFF_Hes | AFF_Str | AFF_Unc | AFF_Foc | AFF_Agg | AFF_Bor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chi-Square | 1,385 | ,308 | 2,096 | 1,159 | ,359 | ,247 | 1,151 | ,924 | ,389 | 1,286 | 11,435 | 1,815 | 6,231 | 3,830 | 5,032 | 4,843 | 9,218 | 1,675 | ,864 |
| df | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Asymp. Sig. | ,500 | ,857 | ,351 | ,560 | ,836 | ,884 | ,562 | ,630 | ,823 | ,526 | ,003 | ,404 | ,044 | ,147 | ,081 | ,089 | ,010 | ,433 | ,649 |

a. Kruskal Wallis Test
b. Grouping Variable: PROFILE

Figure A.5: Kruskal-Wallis test statistics for all the inter-subjects factors

**Ranks**

| | TRAINING | N | Mean Rank |
|---|---|---|---|
| CRED_GLB | 1 | 11 | 13,77 |
| | 2 | 19 | 16,50 |
| | Total | 30 | |
| CRED_AFF | 1 | 11 | 16,50 |
| | 2 | 19 | 14,92 |
| | Total | 30 | |
| CRED_CONF | 1 | 11 | 19,23 |
| | 2 | 19 | 13,34 |
| | Total | 30 | |
| CRED_MOTI | 1 | 11 | 15,32 |
| | 2 | 19 | 15,61 |
| | Total | 30 | |
| CRED_QUAL | 1 | 11 | 16,59 |
| | 2 | 19 | 14,87 |
| | Total | 30 | |
| CRED_TOM | 1 | 11 | 17,00 |
| | 2 | 19 | 14,63 |
| | Total | 30 | |
| DIFF | 1 | 11 | 15,91 |
| | 2 | 19 | 15,26 |
| | Total | 30 | |
| UNDERSTND | 1 | 11 | 16,23 |
| | 2 | 19 | 15,08 |
| | Total | 30 | |
| EMPATHY | 1 | 11 | 18,64 |
| | 2 | 19 | 13,68 |
| | Total | 30 | |
| PLEASANT | 1 | 11 | 16,86 |
| | 2 | 19 | 14,71 |
| | Total | 30 | |
| AFF_TOT | 1 | 11 | 19,00 |
| | 2 | 19 | 13,47 |
| | Total | 30 | |
| AFF_Rel | 1 | 11 | 13,00 |
| | 2 | 19 | 16,95 |
| | Total | 30 | |
| AFF_Emb | 1 | 11 | 19,82 |
| | 2 | 19 | 13,00 |
| | Total | 30 | |
| AFF_Hes | 1 | 11 | 14,14 |
| | 2 | 19 | 16,29 |
| | Total | 30 | |
| AFF_Str | 1 | 11 | 14,82 |
| | 2 | 19 | 15,89 |
| | Total | 30 | |
| AFF_Unc | 1 | 11 | 15,95 |
| | 2 | 19 | 15,24 |
| | Total | 30 | |
| AFF_Foc | 1 | 11 | 20,82 |
| | 2 | 19 | 12,42 |
| | Total | 30 | |
| AFF_Agg | 1 | 11 | 15,91 |
| | 2 | 19 | 15,26 |
| | Total | 30 | |
| AFF_Bor | 1 | 11 | 13,50 |
| | 2 | 19 | 16,66 |
| | Total | 30 | |

**Ranks**

| | XP | N | Mean Rank |
|---|---|---|---|
| CRED_GLB | 1 | 14 | 16,57 |
| | 2 | 16 | 14,56 |
| | Total | 30 | |
| CRED_AFF | 1 | 14 | 16,25 |
| | 2 | 16 | 14,84 |
| | Total | 30 | |
| CRED_CONF | 1 | 14 | 13,79 |
| | 2 | 16 | 17,00 |
| | Total | 30 | |
| CRED_MOTI | 1 | 14 | 17,71 |
| | 2 | 16 | 13,56 |
| | Total | 30 | |
| CRED_QUAL | 1 | 14 | 16,82 |
| | 2 | 16 | 14,34 |
| | Total | 30 | |
| CRED_TOM | 1 | 14 | 16,43 |
| | 2 | 16 | 14,69 |
| | Total | 30 | |
| DIFF | 1 | 14 | 17,29 |
| | 2 | 16 | 13,94 |
| | Total | 30 | |
| UNDERSTND | 1 | 14 | 15,46 |
| | 2 | 16 | 15,53 |
| | Total | 30 | |
| EMPATHY | 1 | 14 | 13,68 |
| | 2 | 16 | 17,09 |
| | Total | 30 | |
| PLEASANT | 1 | 14 | 14,82 |
| | 2 | 16 | 16,09 |
| | Total | 30 | |
| AFF_TOT | 1 | 14 | 13,00 |
| | 2 | 16 | 17,69 |
| | Total | 30 | |
| AFF_Rel | 1 | 14 | 18,07 |
| | 2 | 16 | 13,25 |
| | Total | 30 | |
| AFF_Emb | 1 | 14 | 14,36 |
| | 2 | 16 | 16,50 |
| | Total | 30 | |
| AFF_Hes | 1 | 14 | 15,75 |
| | 2 | 16 | 15,28 |
| | Total | 30 | |
| AFF_Str | 1 | 14 | 13,93 |
| | 2 | 16 | 16,88 |
| | Total | 30 | |
| AFF_Unc | 1 | 14 | 15,71 |
| | 2 | 16 | 15,31 |
| | Total | 30 | |
| AFF_Foc | 1 | 14 | 12,57 |
| | 2 | 16 | 18,06 |
| | Total | 30 | |
| AFF_Agg | 1 | 14 | 16,89 |
| | 2 | 16 | 14,28 |
| | Total | 30 | |
| AFF_Bor | 1 | 14 | 17,29 |
| | 2 | 16 | 13,94 |
| | Total | 30 | |

Figure A.6: Kruskal-Wallis test ranks for the TRAINING and XP factors

**Ranks**

| | TOM | N | Mean Rank |
|---|---|---|---|
| CRED_GLB | 1 | 15 | 15,97 |
| | 2 | 15 | 15,03 |
| | Total | 30 | |
| CRED_AFF | 1 | 15 | 15,43 |
| | 2 | 15 | 15,57 |
| | Total | 30 | |
| CRED_CONF | 1 | 15 | 15,13 |
| | 2 | 15 | 15,87 |
| | Total | 30 | |
| CRED_MOTI | 1 | 15 | 17,17 |
| | 2 | 15 | 13,83 |
| | Total | 30 | |
| CRED_QUAL | 1 | 15 | 16,30 |
| | 2 | 15 | 14,70 |
| | Total | 30 | |
| CRED_TOM | 1 | 15 | 16,67 |
| | 2 | 15 | 14,33 |
| | Total | 30 | |
| DIFF | 1 | 15 | 14,93 |
| | 2 | 15 | 16,07 |
| | Total | 30 | |
| UNDERSTND | 1 | 15 | 15,23 |
| | 2 | 15 | 15,77 |
| | Total | 30 | |
| EMPATHY | 1 | 15 | 16,07 |
| | 2 | 15 | 14,93 |
| | Total | 30 | |
| PLEASANT | 1 | 15 | 16,70 |
| | 2 | 15 | 14,30 |
| | Total | 30 | |
| AFF_TOT | 1 | 15 | 15,67 |
| | 2 | 15 | 15,33 |
| | Total | 30 | |
| AFF_Rel | 1 | 15 | 16,90 |
| | 2 | 15 | 14,10 |
| | Total | 30 | |
| AFF_Emb | 1 | 15 | 13,93 |
| | 2 | 15 | 17,07 |
| | Total | 30 | |
| AFF_Hes | 1 | 15 | 14,80 |
| | 2 | 15 | 16,20 |
| | Total | 30 | |
| AFF_Str | 1 | 15 | 16,33 |
| | 2 | 15 | 14,67 |
| | Total | 30 | |
| AFF_Unc | 1 | 15 | 13,97 |
| | 2 | 15 | 17,03 |
| | Total | 30 | |
| AFF_Foc | 1 | 15 | 15,07 |
| | 2 | 15 | 15,93 |
| | Total | 30 | |
| AFF_Agg | 1 | 15 | 13,60 |
| | 2 | 15 | 17,40 |
| | Total | 30 | |
| AFF_Bor | 1 | 15 | 13,83 |
| | 2 | 15 | 17,17 |
| | Total | 30 | |

**Ranks**

| | PROFILE | N | Mean Rank |
|---|---|---|---|
| CRED_GLB | 1 | 10 | 13,10 |
| | 2 | 10 | 17,10 |
| | 3 | 10 | 16,30 |
| | Total | 30 | |
| CRED_AFF | 1 | 10 | 14,35 |
| | 2 | 10 | 16,35 |
| | 3 | 10 | 15,80 |
| | Total | 30 | |
| CRED_CONF | 1 | 10 | 18,40 |
| | 2 | 10 | 14,85 |
| | 3 | 10 | 13,25 |
| | Total | 30 | |
| CRED_MOTI | 1 | 10 | 17,05 |
| | 2 | 10 | 16,20 |
| | 3 | 10 | 13,25 |
| | Total | 30 | |
| CRED_QUAL | 1 | 10 | 15,65 |
| | 2 | 10 | 14,30 |
| | 3 | 10 | 16,55 |
| | Total | 30 | |
| CRED_TOM | 1 | 10 | 16,35 |
| | 2 | 10 | 15,70 |
| | 3 | 10 | 14,45 |
| | Total | 30 | |
| DIFF | 1 | 10 | 13,45 |
| | 2 | 10 | 15,55 |
| | 3 | 10 | 17,50 |
| | Total | 30 | |
| UNDERSTND | 1 | 10 | 13,50 |
| | 2 | 10 | 15,95 |
| | 3 | 10 | 17,05 |
| | Total | 30 | |
| EMPATHY | 1 | 10 | 15,95 |
| | 2 | 10 | 16,35 |
| | 3 | 10 | 14,20 |
| | Total | 30 | |
| PLEASANT | 1 | 10 | 16,55 |
| | 2 | 10 | 13,05 |
| | 3 | 10 | 16,90 |
| | Total | 30 | |
| AFF_TOT | 1 | 10 | 17,40 |
| | 2 | 10 | 8,10 |
| | 3 | 10 | 21,00 |
| | Total | 30 | |
| AFF_Rel | 1 | 10 | 12,80 |
| | 2 | 10 | 15,60 |
| | 3 | 10 | 18,10 |
| | Total | 30 | |
| AFF_Emb | 1 | 10 | 19,70 |
| | 2 | 10 | 10,10 |
| | 3 | 10 | 16,70 |
| | Total | 30 | |
| AFF_Hes | 1 | 10 | 12,40 |
| | 2 | 10 | 14,30 |
| | 3 | 10 | 19,80 |
| | Total | 30 | |
| AFF_Str | 1 | 10 | 15,10 |
| | 2 | 10 | 11,30 |
| | 3 | 10 | 20,10 |
| | Total | 30 | |
| AFF_Unc | 1 | 10 | 14,95 |
| | 2 | 10 | 11,55 |
| | 3 | 10 | 20,00 |
| | Total | 30 | |
| AFF_Foc | 1 | 10 | 19,00 |
| | 2 | 10 | 8,60 |
| | 3 | 10 | 18,90 |
| | Total | 30 | |
| AFF_Agg | 1 | 10 | 16,65 |
| | 2 | 10 | 12,80 |
| | 3 | 10 | 17,05 |
| | Total | 30 | |
| AFF_Bor | 1 | 10 | 17,40 |
| | 2 | 10 | 15,10 |
| | 3 | 10 | 14,00 |
| | Total | 30 | |

Figure A.7: Kruskal-Wallis test ranks for the TOM and PROFILE factors

**Test Statistics[b]**

| | AFF_Emb | AFF_Foc | CRED_CONF |
|---|---|---|---|
| Mann-Whitney U | 57,000 | 46,000 | 63,500 |
| Wilcoxon W | 247,000 | 236,000 | 253,500 |
| Z | -2,045 | -2,518 | -1,908 |
| Asymp. Sig. (2-tailed) | ,041 | ,012 | ,056 |
| Exact Sig. [2*(1-tailed Sig.)] | ,042[a] | ,011[a] | ,077[a] |

a. Not corrected for ties.
b. Grouping Variable: TRAINING

**Ranks**

| | TRAINING | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| AFF_Emb | 1 | 11 | 19,82 | 218,00 |
| | 2 | 19 | 13,00 | 247,00 |
| | Total | 30 | | |
| AFF_Foc | 1 | 11 | 20,82 | 229,00 |
| | 2 | 19 | 12,42 | 236,00 |
| | Total | 30 | | |
| CRED_CONF | 1 | 11 | 19,23 | 211,50 |
| | 2 | 19 | 13,34 | 253,50 |
| | Total | 30 | | |

Figure A.8: Mann-Whitney test statistics for the TRAINING factors

**Test Statistics[b]**

| | AFF_TOT | AFF_Emb | AFF_Foc |
|---|---|---|---|
| Mann-Whitney U | 20,000 | 20,000 | 21,000 |
| Wilcoxon W | 75,000 | 75,000 | 76,000 |
| Z | -2,268 | -2,272 | -2,193 |
| Asymp. Sig. (2-tailed) | ,023 | ,023 | ,028 |
| Exact Sig. [2*(1-tailed Sig.)] | ,023[a] | ,023[a] | ,029[a] |

a. Not corrected for ties.
b. Grouping Variable: PROFILE

**Test Statistics[b]**

| | AFF_TOT | AFF_Emb | AFF_Foc |
|---|---|---|---|
| Mann-Whitney U | 6,000 | 26,000 | 10,000 |
| Wilcoxon W | 61,000 | 81,000 | 65,000 |
| Z | -3,326 | -1,816 | -3,025 |
| Asymp. Sig. (2-tailed) | ,001 | ,069 | ,002 |
| Exact Sig. [2*(1-tailed Sig.)] | ,000[a] | ,075[a] | ,002[a] |

a. Not corrected for ties.
b. Grouping Variable: PROFILE

**Test Statistics[b]**

| | AFF_TOT | AFF_Emb | AFF_Foc |
|---|---|---|---|
| Mann-Whitney U | 39,000 | 38,000 | 44,000 |
| Wilcoxon W | 94,000 | 93,000 | 99,000 |
| Z | -,832 | -,907 | -,454 |
| Asymp. Sig. (2-tailed) | ,406 | ,364 | ,650 |
| Exact Sig. [2*(1-tailed Sig.)] | ,436[a] | ,393[a] | ,684[a] |

a. Not corrected for ties.
b. Grouping Variable: PROFILE

**Ranks**

| | PROFILE | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| AFF_TOT | 1 | 10 | 13,50 | 135,00 |
| | 2 | 10 | 7,50 | 75,00 |
| | Total | 20 | | |
| AFF_Emb | 1 | 10 | 13,50 | 135,00 |
| | 2 | 10 | 7,50 | 75,00 |
| | Total | 20 | | |
| AFF_Foc | 1 | 10 | 13,40 | 134,00 |
| | 2 | 10 | 7,60 | 76,00 |
| | Total | 20 | | |

**Ranks**

| | PROFILE | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| AFF_TOT | 2 | 10 | 6,10 | 61,00 |
| | 3 | 10 | 14,90 | 149,00 |
| | Total | 20 | | |
| AFF_Emb | 2 | 10 | 8,10 | 81,00 |
| | 3 | 10 | 12,90 | 129,00 |
| | Total | 20 | | |
| AFF_Foc | 2 | 10 | 6,50 | 65,00 |
| | 3 | 10 | 14,50 | 145,00 |
| | Total | 20 | | |

**Ranks**

| | PROFILE | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| AFF_TOT | 1 | 10 | 9,40 | 94,00 |
| | 3 | 10 | 11,60 | 116,00 |
| | Total | 20 | | |
| AFF_Emb | 1 | 10 | 11,70 | 117,00 |
| | 3 | 10 | 9,30 | 93,00 |
| | Total | 20 | | |
| AFF_Foc | 1 | 10 | 11,10 | 111,00 |
| | 3 | 10 | 9,90 | 99,00 |
| | Total | 20 | | |

Figure A.9: Mann-Whitney test statistics for the PROFILE factors

# Appendix B

# TARDISx evaluation questionnaire

This appendix presents the 2-pages questionnaire, in French, used for both TARDISx experiment's background survey and post-hoc evaluation. The participants were asked not to look at the post-hoc questions until the job interviews were done.