Auton Agent Multi-Agent Syst (2010) 20:401–420
DOI 10.1007/s10458-009-9083-z

Coping with shame and sense of guilt: a Dynamic Logic Account

Paolo Turrini · John-Jules Ch. Meyer · Cristiano Castelfranchi

Published online: 29 April 2009

Abstract Aim of this work is to provide a formal characterization of those emotions that deal with normative reasoning, such as shame and sense of guilt, to understand their relation with rational action and to ground their formalization on a cognitive science perspective. In order to do this we need to identify the factors that constitute the preconditions and trigger the reactions of shame and sense of guilt in cognitive agents, that is *when* agents feel ashamed or guilty and *what* agents do when they feel so. We will also investigate how agents can induce and silence these feelings in themselves, i.e. the analysis of defensive strategies they can employ. We will argue that agents do have control over their emotions and we will analyze some operations they can carry out on them.

Keywords Emotions · Logical foundations of AI · Agent theory

> "You know, everybody makes mistakes when they are president"
> (William Jefferson Clinton)

1 Introduction

In January 1998, the President of US pronounces one of his most famous statements:

> I did not have sexual relations with that woman, Miss Lewinsky. I never told anybody to lie, not a single time - never. These allegations are false. And I need to go back to work for the American people.

In August 1998, after evidence about the Monica Lewinsky sex scandal is released, he amends it:

P. Turrini (✉) · J.-J. Ch. Meyer
University of Utrecht, Utrecht, The Netherlands
e-mail: paolo@cs.uu.nl

C. Castelfranchi
ISTC-CNR, Rome, Italy

> I did have a relationship with Ms Lewinsky that was not appropriate. In fact, it was wrong. It constituted a critical lapse in judgment and a personal failure on my part for which I am solely and completely responsible.

(Source: [9])

If we ask ourselves why his first reaction was to deny he did something wrong, we could agree with his explanation: "a desire to protect myself from the embarrassment of my own conduct." But why did the president of US allow the American media reporting details of his private life while in a first moment he said that "he will refuse to answer detailed questions because of privacy considerations affecting his family and in an effort to preserve the dignity of his office" [5]? It seems that in the struggle between the right to have a private life and right for the Public Opinion to know, Clinton chose eventually the latter, exposing him to what media called his "days of shame" [9,5].

Aim of this work is to provide a formal characterization of those emotions that deal with normative reasoning, such as shame and sense of guilt, to understand their relation with rational action and to ground their formalization on a cognitive science perspective. In order to do this we need to identify the factors that constitute the preconditions and trigger the reactions of shame and sense of guilt in cognitive agents, that is *when* agents feel ashamed or guilty and *what* agents do when they feel so. We will also investigate how agents can induce and silence these feelings in themselves, i.e. the analysis of defensive strategies they can employ. We will argue that agents do have control over their emotions and we will analyze some operations they can carry out on them. We will maintain moral judgment to be socially determined (by the so called Significant Others [31]) and we will see that many are the ways of manipulating one's own emotions depending on one's psychological inclination. Coming back to our example, we will see what type of agent the president of the US has been and what else he could have done in order not to feel embarrassed or responsible.

1.1 Related work

As witnessed by [14] the study of emotions has recently gained much attention in the fields of Artificial Intelligence [29,19], Evolutionary Computation [30] and Multi Agent Systems [26], due to the encounter between computer science tools and neuro, cognitive and social sciences analyses [11,24,23]. Ours is a cognitive perspective: even though we agree that it is important to study emotions from a computational and emergentist point of view, we argue that in order to build an *anatomy of emotions* ([7]) it is as important to understand them in terms of their interaction with other cognitive ingredients.[1]

The most influential cognitive paradigm for studying and constructing cognitive agents with emotions has been that by Ortony, Clore and Collins [24]. Nevertheless, in [24] the characterizations of feelings related to norms are not deeply investigated:

[1] Many criticisms have been moved to cognitivists theories, some of which can be hardly addressed in this context. Nevertheless we would like to point out how several ones are based on what we think is a misconception of the use of formal models of cognition. In the study of normative emotions of [30] it is argued that "logic does not provide an adequate foundation" to the study of human behaviour and "the necessary abandonment of logical models for the explanation and simulation of human social behaviour" is advocated. Even though we share the worries in [30] with respect to representing humans as perfect reasoners, we claim that an anti-logical position in modelling interaction is simply wrong: emotions can be studied as mechanisms that act on human cognition. But mechanisms do have a logic. What is more, the recent breakthroughs of logical models in the study of social interaction and information flow [34] have shown that formal semantics can lead to the construction of rigorous models of complex phenomena such as emotions (as in [19]).

> "In order to feel shame one must have violated a standard one takes to be important, as moral standards are. Such violations are held to be inexcusable. This is not necessary for a person who is feeling guilty.(...) In fact, we do not think that there is a distinct emotion of feeling guilty. Rather, we view feelings of guilt as mixtures of distinct emotions such as shame and regret, perhaps accompanied by certain cognitive states, such as the belief that one was, at least technically, responsible." (p. 142–143)

Many expressions here would need to be explicated further: why are violations only in case of shame held to be inexcusable? What is a mixture of emotions? And a technical responsibility? If we find the distinction between shame and sense of guilt and all the other related feelings as meaningful at all, we need to have clear-cut definitions that relate those feelings to agents' mental states and to precisely understand their functioning.

We will pursue a formal investigation on emotions, as done for instance in [26], but adding a closer look to the formal properties of our notions, that we construct within a well known logical framework such as KARO [36,37,19]. This means that unlike [26] we will provide a formal semantics to our language, that will allow to derive logical properties. Formal models of emotional agents are used in field related to Multi Agent Systems, such as Game Theory [4]. But logical agents are not only abstract entities. The idea of implementing them is well rooted in computer science and it dates back to Shoham's AGENT-0 [28]. A cognitive formal approach in programming agents with emotions is also adopted by [12], in which a logic-based agent-oriented programming language is devised, which is inspired by 3APL [13], and that can be used to implement emotional agents. We share with [12] the logical framework on which we ground the construction of emotional states. Nevertheless we will not focus on the properties of the agents programs, but on the dynamics of emotional agents. From a cognitive point of view we will follow the analysis of [22] that grounds emotional displays like blushing and feelings like loneliness or pride on complex Multi Agent interaction. We claim that such model overcomes the oversimplifications in [24] and it is grounded on clear intuitions that ease a proper formal investigation.

Structure of the paper The paper is structured as follows: We will first present a cognitive model of shame and sense of guilt as social emotions, isolating the preconditions and the reactions that accompany them. We will then provide a formal representation of shame, sense of guilt and their dynamics in terms of basic notions such as beliefs, goals and violations, following the rational action approach of [36,37,19,32]. We will analyse the language providing connections with the Propositional Dynamic Logic proposed in [17] and extensively used in Computer Science, extending it to treat Multi Agent Systems and Collective Actions.

2 A cognitive model for shame and sense of guilt

In this section we are going to discuss the cognitive model developed by Castelfranchi, Miceli and Poggi in [22,8,7], that takes into account the Multi Agent nature of emotions such as shame and sense of guilt (described as *social emotions*), linking them to the notions of agency and norm violation.

The belief of responsibility Miceli and Castelfranchi [22] emphasize how the perceived causal responsibility [10] and [33], that is the belief of having had the capacity to avoid a damage or a violation, is a crucial notion for distinguishing these feelings: "when ashamed,

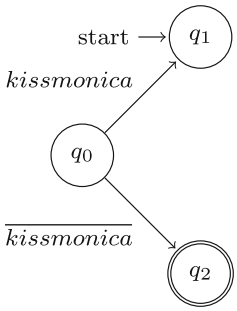


Fig. 1 Epistemic state of agents feeling guilty

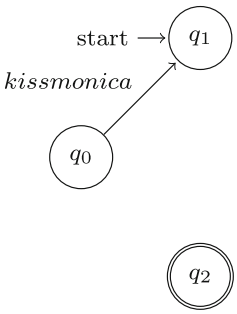


Fig. 2 Epistemic state of agents feeling ashamed

one sees oneself as incompetent or inadequate with respect to some goal; when guilty, one sees oneself as endowed with negative power." ([22], p. 295).[2]

In a nutshell we can focus on the following statements:

– In case of sense of guilt, the agent (say Bill) believes that he could have avoided what he did. Let us suppose that in this past moment he had an action $kissmonica$ that lead to a bad state, and an alternative action $\overline{kissmonica}$ that lead to a good state. And that he chose the first.
In the Fig. 1 the $PDL^{-,\neg}$ formula[3] $\langle kissmonica\rangle^{-1}\langle\overline{kissmonica}\rangle OK$ is true at q_1, where OK is the atom made true at the double-circled states. The whole sentence means: "If he had not done $kissmonica$ he could have done $\overline{kissmonica}$ and this would have avoided a damage".

– In case of shame, Bill believes he could not avoid the present bad situation. In the Fig. 2, $\langle kissmonica\rangle^{-1}\langle\overline{kissmonica}\rangle OK$ is false at q_1. The whole sentence means "If he had not done $kissmonica$ he could not have done any other action to avoid the damage". In this case Bill is so much in love with Monica that he just cannot avoid kissing her.

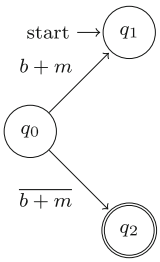
It seems fundamental to represent actions, ability to refrain, and belief about the consequences of them, with respect to a given set of deontic statements.

In a Multi Agent System we can moreover imagine situations in which a bad state is avoided only by two agents working together. For instance Bill and Monica could avoid to have a relation (see Fig. 3).

[2] In Maria Miceli's words shame is the perception of oneself being a dull knife, whereas feeling guilty is the perception of oneself being a sharp knife (Maria Miceli "Personal Communication").

[3] We call $PDL^{-,\neg}$ the variant of Propositional Dynamic Logic [17] with converse operator and atomic action negation.

Fig. 3 Collective guilt



The Significant Others Following [31], we conceive of the notion of misbehaviour as having a social connotation. According to this account, deontic judgments are socially determined. Agents import such judgments from a particular subset of other agents, those ones on which they are dependent, they care about, and whose standards they have the goal to meet: the *Significant Others*. Significant Others are important for the decision of accepting or not accepting a deontic judgment. When agents use values or norms as input for their decisions to comply with what is prescribed, we say that these values or norms have been internalized. An internalized norm, thus, drives the agents' behaviour towards obedience. Psychological research has shown that internalization can be induced, by Significant Others' expectations [31] and commands [15]. In this respect it is interesting to consider the reaction to the Lewinski case of the former German chancellor Helmut Kohl. In an interview with the German newspaper Die Welt, the chancellor said it was not up to him to pass judgement on Mr. Clinton's private life. But he said the way people all over the world were following the most private details on the Internet was enough to make him sick [5]. Kohl points out how "people all over the world" were judging some facts that are usually "private details". As we already observed, in the Lewinski case Public Opinion was left to determine which behaviours were acceptable and which were not (being faithful, say the truth, let other people believe in his being faithful, etc.).

In this article the dynamics of internalization will be taken into account with respect to the way changing Significant Others influences the set of world states agents label as good and acceptable, and what the consequences of these are in terms of feeling guilty and ashamed. For instance keeping American people ignorant or not concerned with his personal affairs would have allowed President Clinton to avoid public exposure. But different agents may have different strategies to cope with their emotions [31]. We may indeed observe, due to difference in personality traits, a pride reaction in some agents, while instead others may show feelings of loneliness and exclusion (Maria Miceli "Personal Communication"). We can imagine high self esteem agents that tend to pride, while low self esteem that tend to show loneliness. Personal feelings can also come up, like sense of guilt and shame towards the self. The interplay is sketched in Table 1.

Table 1 Emotional reactions in a Multi Agent System

	Others do not share		Others share	
	Internalized norm/value	Not internalized norm/value	Internalized norm/value	Not internalized norm/value
Negative power	Sense of guilt towards the self	Indifference	Sense of guilt	Pride/Loneliness
Inadequacy	Shame towards the self	Indifference	Shame	Pride/Loneliness

We can notice that agents can have fail to accomplish values they personally have, this is the reason why guilt and shame can be addressed towards the self. We believe that the different opinions that an agent and its Significant Others have concerning a relevant value can be associated with phenomena like cognitive dissonance and the processes that lead to its resolution [15].

2.1 Controlling emotions

One of the most interesting features of the cognitive approach in [22,8,7] is that of human beings having some control of their emotions. "In particular, people can react to their own emotions defending themselves from the disturbing or blamed ones. They try to repress, deny, and manipulate them." [7].

We are going to develop this intuition by describing both *control via belief manipulation* and *control via goal manipulation*. The first type of control will act on an agent belief base in order to eliminate those beliefs that induce the emotion in the self, while the second will act with the same purpose on the agent goal base. But how would these strategies work? Elaborating on Miceli and Castelfranchi's proposal in [22], the different types of reaction can be related to basic psychological types of agents, such as high self esteem (HSE) and low self esteem (LSE) ones. HSE people might try to question the basis on the grounds of which they are blamed or they feel bad; they will react actively by trying to find justifications for their actions. Instead LSE agents will tend to apologise and find excuses, as they are more unlikely to question the values which they are accused to go against. In this sense HSE agents will tend to react with pride to the above mentioned situation, while LSE agents will react with for instance feeling lonely or rejected. We distinguish a typical LSE reaction, "I did not know", that we classify as an excuse by claiming ignorance of the relevant effects of a dangerous action that has been carried out, from a typical HSE reaction, "It was not that bad", that is a justification on a presumed violation. It claims that what others may think as wrong is not really so.

Among the beliefs agents can control is indeed that of an agent being a Significant Other. This last type of control is of fundamental importance, because it allows to modify the moral valuations agents have about the world. What we mean here is the following: Significant Others have the role to tell what it is important/good for an agent to do, and in this sense modifying Significant Others means to automatically change evaluation of what is good or bad.

We argue that these mechanisms can be formally modeled.

3 Emotion dynamics in shame and sense of guilt

In this part we analyze shame and sense of guilt in terms of their preconditions and postconditions, that is what are the cognitive states that cause someone to feel ashamed or guilty and what are the possible reactions that such emotions trigger.

3.1 Shame

Preconditions As argued in [7] feeling ashamed always involves a believed negative self evaluation (concerning one's inadequacy) related to somebody whose judgment agents care about. In this sense a precondition for feeling ashamed is *the belief of not having had a capacity to get over a bad state*.

Reactions As far as the behaviour is concerned, the goal of an ashamed person is to reduce exposure [8]. This will be translated into various actions, either *minimizing the importance of a value* (agents that are proud of something), which is an HSE agent's typical reaction, or *minimizing their active contribution to a damage* by declaring submission and imperfection, which are typical LSE agents' reactions.[4]

Clinton's reaction was clearly an LSE reaction: he firstly did not admit that what he did was wrong (it was an extreme case of minimization of one's active contribution to a damage), and when he admitted it he never justified himself. One alternative answer could be summarized by the exclamation "It's not your business!"—a clear HSE reaction—holding others not concerned with the wrongdoing.

3.2 Feeling guilty

Preconditions The feeling of guilt is usually linked to the conviction of having actively injured someone or broken some moral imperative or norm [7]. It will be associated to the *evaluation of having negative power against a given situation*, i.e. a perception of responsibility. From agent types reactions and beliefs of responsibility, it is intuitively clear how difficult it is to both feel ashamed and guilty at the same time for the very same thing. Either agents believe to be a dull knife or a sharp knife, i.e. holding consistent beliefs either does an agent believe he actively made a mistake (while being able to avoid it) or he beliefs he could not have avoided it. Nevertheless the transitions between the two feelings can be carried out by means of a belief revision [36] concerning responsibility.

Reactions Reactions in case of sense of guilt are similar to that of shame for HSE and LSE agents. But there are some more we mention for the sake of completeness. As far as reactive behaviour is concerned, one first goal of agents that feel guilty is that of reparation, which triggers the agent to care about the damaged person, and to expiate, to pay in some sense for what has been done. A very interesting property of agents that feel guilty is to *regret doing something*, that act as a situation marker [11]. Regret can be exemplified by the sentences "I wish I did not do it". We can thus say that regret is the desire of not having done something that was actually done. Reactions in the direction of *cancelling the belief of ability or of importance of the action made* will be modeled.

4 A language for rational agents with emotions

4.1 Emotional Multi Agent KARO

The reconstruction of cognitive agents that feel guilty and ashamed needs some further concepts with respect to those already in [19]. We need to reason about agents' deontic judgments, their mismatch with what they believe others have and operations that act directly on the agent emotional state, i.e. allow changing perception of situations in order not to feel bad. The work of Meyer in Dynamic Deontic Logic [18] and of Meyer and colleagues in the logic of emotions [19] provide a solid basis for addressing this issue.

What we need to do more is to extend the formalism in order to capture the Multi Agent aspects of actions and emotions, how for instance many agents can feel guilty for something

[4] In [8] it is argued how emotional displays of shame are admissions of imperfection, for instance blushing. Blushing is not a confession of guilt but it still has a precise communicative function. We can be caught doing something that looks bad with respect to others (even though we know they are wrong) and yet blush.

they independently or collectively caused. The analysis of collective responsibility in organizational and Multi Agent settings has been addressed for instance in [16,33] and logics for Multi Agent interaction flourish [2,25]. To our knowledge no attempt has so far been made to bridge the gap between Multi Agent interaction research and the formal treatment of emotions.

Our plan is then to start with a Multi Agent extension of the KARO framework. We will introduce in such a framework a set of violation constants V_i indexed with agents, that will label worlds that are bad for particular agents, as well as different dynamics for different types of agents. Even though we will in fact be able to talk about more agents in this framework, collective actions will be only addressed in the last part.

4.2 The language

The actions that agents perform can have different forms. We classically consider a finite set of atomic actions from which all the others can be obtained compositionally. We will not consider parallel execution of action—that will be instead used in the last part to talk about collective action—while action negation (i.e. refraining) will be limited to the atomic case.[5] Finally the planning component in emotions requires the treatment of the notion of action composition and iteration. The set of action expressions Act is the smallest set containing all actions of the following form:

$$\alpha ::= b|\mathbf{eliminate}(j)|\mathbf{welcome}(j)|\alpha;\alpha|\alpha^*$$

where $b = a|\overline{a}$ is an atomic action or its negation; $i, j \in Agt$, which is the set of agents; $eliminate$, $welcome$ actions will be used for updating evaluations and will be dealt with as special actions later on in the paper. The informal meaning of these actions is that an agent changes its evaluation of the world states, by eliminating or adding Significant Others. The operation of union, composition ; and iteration $*$ are standard regular operations [17]. The set of events Evt[6] has the following grammar:

$$\xi ::= i : \alpha|\xi \cup \xi|\xi;\xi|\xi^*$$

The grammar of events language is computationally quite simple. We restrict ourselves to such grammar for the sake of simplicity. More complex extensions are possible, for instance talking about complex plans or group actions or even parallel group actions. We point out that events are more than just actions. In classical dynamic logic there is no notion of agency in a Multi Agent context. By introducing the notion of event, we will allow for agents being associated to a certain action.[7]

To model transitions we will use the KARO metaphor of action as model update [37]. The idea is that actions make us jump to new worlds in which the original relations are changed. Say if, in a situation w where we do not know whether p, we learn that p holds, we end up in a new model where the epistemic possibilities stemming from w will agree with p.

[5] For the technical reasons see [18].

[6] As it will be clear from the semantics, what we call here *events* is elsewhere understood as action. We prefer to stick to the word event as done in Dynamic Epistemic Logic, Product Update [3] and Agent Based Logic Programming [12]. The reader has to be aware that this is not the only possible interpretation of such concept.

[7] In our framework a Single Agent System can be obtained just dropping the agent index i from the event semantics, that becomes $\xi ::= \alpha|\xi \cup \xi|\xi;\xi|\xi^*$. In a Single Agent System events are reducible to actions.

On the Multi Agent structure of events We would like to work on models that allow for events to be indeterministic in the future, while keeping a linear past. The reader can think of a history as an intuitive model for the transitions we have in mind. In a history every node has a unique predecessor (the course events took) while allowing for multiple successors (the course events may take).

Formally we impose an order < that links pairs model-world with events.

$$<= \{(\langle M, w\rangle, \langle M', w'\rangle) | \exists \xi \in Evt \text{ s.t. } \langle M', w'\rangle \in [\![\xi]\!]_R \langle M, w\rangle\}.$$

For all $\langle M, w\rangle$, there is a unique $\langle M', w'\rangle$ such that $\langle M', w'\rangle < \langle M, w\rangle$. Moreover we impose that this predecessor is reachable by a unique transition. $\langle M', w'\rangle \in [\![\xi]\!]_R \langle M, w\rangle$ implies that for all $\xi' \neq \xi$ not $\langle M', w'\rangle \in [\![\xi']\!]_R \langle M, w\rangle$.

This guarantees linear past and branching future and it has a syntactic counterpart that is expressible in the logic of programs [17].

4.2.1 Syntax

We assume a finite set of agents Agt and a countable set of atomic propositions Π_0. Moreover we introduce special atoms V_i for $i \in Agt$. Special atoms are extensively used to describe also emotional states and agent types. Our language is given by the following syntax:

$$\phi ::= p | L(i) | H(i) | guilty(i, a, j) | ashamed(i, a, j) | V_i | \neg\phi | \phi \wedge \psi | \\ \mathbf{Sig}_{i,j} | \mathbf{B}_i\phi | \mathbf{D}_i\phi | [\xi]\phi | [\xi]^{-1}\phi$$

where $p \in \Pi_0$ (the set of atomic propositions) , $i, j \in Agt, \alpha \in Act, \xi \in Evt$. $L(i), H(i)$ indicate a Low and High Self Esteem personality type of the agent, $guilty(i, a, j)$ and $ashamed(i, a, j)$ refers to an agent i feeling guilty or ashamed for an action a relative to an other agent j. They will be later defined as abbreviations. The informal reading of the modalities is "i is a Significant Other for j", "i believes that ϕ is true", "i desires that ϕ is true" , "after ξ, ϕ becomes true", "before ξ, ϕ was true".

We moreover use the following abbreviations: $\phi \vee \psi := \neg(\neg\phi \wedge \neg\psi)$; $\phi \rightarrow \psi := \neg\phi \vee \psi$; $\phi \leftrightarrow \psi := (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$; $\langle\xi\rangle\phi := \neg[\xi]\neg\phi$, which are usual in modal logic.

In our logic we add the following abbreviations: $\mathbf{P}\phi =: \langle(\bigvee_{i\in Agt} \bigvee_{a\in Act}(i : a))\rangle^{-1}\phi$ (ϕ was true just before the latest action), $\mathbf{DONE}_j(b) := \langle(j : b)\rangle^{-1}(p \vee \neg p)$ (Agent j did b), $\mathbf{DONE}_j(\overline{b}) := \bigvee_{c\in Act, c\neq b}\langle(j : c)\rangle^{-1}(p \vee \neg p)$ (Agent j did not do b).

4.2.2 Structure

The structures interpreting the language $\mathcal{L}$ are given by a class of E-KAROUS[8] models in which each model M is of the following shape:

$$M = \langle Agt, W, Act, \sigma, \{B_i | i \in Agt\}, \{D_i | i \in Agt\}, Aut, \rangle$$

In which Agt is a finite nonempty set of agents. We take $Agt = Agt_h \cup Agt_l$. We require $Agt_h \cap Agt_l = \emptyset$: the set of agents is partitioned into Agt_h which will represent the HSE agents and Agt_l, the LSE agents. W is a nonempty set of worlds; Act is the finite set of basic actions;

[8] E-KAROUS is a fancy transformation of KARO to an emotional Multi Agent shape, resembling moreover the name of a person that did not pay much attention to the suggestions of his friends.

$$\sigma : \Pi_0 \cup \bigcup_{i,j \in Agt} \bigcup_{a \in Act} guilty(i, a, j) \cup \bigcup_{i,j \in Agt} \bigcup_{a \in Act} ashamed(i, a, j) \rightarrow 2^W$$

is the augmented valuation function, that assigns each atom to a set of worlds, with the intended meaning that they are those worlds in which the atom is true. $\{B_i | i \in Agt\}$ is an epistemic accessibility relation; Each $B_i \subseteq W \times W$ is composed by pairs $\{w, w'\}$ in such a way that the world w' represents an epistemic alternative for agent i at world w. We indicate with $[w]_{B_i}$ the set of epistemic alternatives for agent i at world w. $\{D_i | i \in Agt\}$ is defined as B_i for desired worlds. Aut is a function g such that $g : Agt \times Agt \times W \rightarrow 2^W$. This function associates to pair of agents and a situation a set of situations. $\{v\} = Aut(i, j, w)$ can be then read as "In world w, world v is the only world considered bad by agent i, because of agent j". We may have more agents that make someone dislike some situation. We say that those agents *block* that situation. We label as $Sig_{(i,w)} = \{j | \exists w' s.t. w' \in Aut(i, j, \langle w \rangle)\}$ the set of Significant Others for agent i at the situation $\langle M, w \rangle$. This set comprises those agents that block at least one situation in a given situation for agent i.

4.2.3 Semantics

The formulas of our language $\mathcal{L}$ are interpreted as follows:

- $M, w \models p$ iff $w \in \sigma(p)$; Propositional cases are treated as usual;
- $M, w \models L(i)$ iff $i \in Agt_l$;
- $M, w \models H(i)$ iff $i \in Agt_h$;
- $M, w \models guilty(i, a, j)$ iff $w \in \sigma(guilty(i, a, j))$
- $M, w \models ashamed(i, a, j)$ iff $w \in \sigma(ashamed(i, a, j))$
- $M, w \models V_i$ iff $w \in \bigcup_j Aut(i, j, w)$;
- $M, w \models \mathbf{B}_i \phi$ iff $M, w' \models \phi$ for all w' s.t. $w B_i w'$;
- $M, w \models \mathbf{D}_i \phi$ iff $M, w' \models \phi$ for all w' s.t. $w D_i w'$;
- $M, w \models [\xi]\phi$ iff $M', w' \models \phi$ for all $\langle M', w' \rangle$ s.t. $\langle M, w \rangle [\![\xi]\!]_R \langle M', w' \rangle$;
- $M, w \models [\xi]^{-1}\phi$ iff $M', w' \models \phi$ for all $\langle M', w' \rangle$ s.t. $\langle M', w' \rangle [\![\xi]\!]_R \langle M, w \rangle$;
- $M, w \models \mathbf{Sig}_{i,j}$ iff $j \in Sig_{(i,w)}$;

where $[\![\,]\!]_R$ is a relation between models, that accounts for the change in the situations brought about by actions. Its semantics and behaviour are analyzed in depth in the Appendix.

4.2.4 Constraints on the models

We denote with $[\![\phi]\!]_M$ the set A such that $A = \{w | M, w \models \phi\}$. We constrain our models in the following way:

- For all $w \in W, [w]_{B_i} \neq \emptyset$;
- $w' \in [w]_{B_i} \Rightarrow [w']_{B_i} = [w]_{B_i}$
- $[\![\mathbf{Sig}_{i,j}]\!]_M \subseteq [\![\mathbf{B}_j \mathbf{Sig}_{i,j}]\!]_M$;
- $W \setminus [\![\mathbf{Sig}_{i,j}]\!]_M \subseteq [\![\mathbf{B}_j \neg \mathbf{Sig}_{i,j}]\!]_M$;
- $[\![V_i]\!]_M \subseteq [\![\mathbf{B}_i V_i]\!]_M$;
- $W \setminus [\![V_i]\!]_M \subseteq [\![\mathbf{B}_i \neg V_i]\!]_M$

Notice that the first two properties force the belief modality to be serial, transitive and euclidean.

4.2.5 Validities

We recall that $M \models \phi$ indicates that $M, w \models \phi$, for any world w; and that $\mathcal{M} \models \phi$ if $M \models \phi$ for any M in the class of models $\mathcal{M}$.

In the class of E-KAROUS models $\mathcal{M}$ this holds:

- $\mathcal{M} \models \mathbf{B}_i \top$;
- $\mathcal{M} \models \neg \mathbf{B}_i \phi \rightarrow \mathbf{B}_i \neg \mathbf{B}_i \phi$;
- $\mathcal{M} \models \mathbf{B}_i \phi \rightarrow \mathbf{B}_i \mathbf{B}_i \phi$;
- $\mathcal{M} \models \mathbf{Sig}_{i,j} \rightarrow \mathbf{B}_i \mathbf{Sig}_{i,j}$;
- $\mathcal{M} \models \neg \mathbf{Sig}_{i,j} \rightarrow \mathbf{B}_i \neg \mathbf{Sig}_{i,j}$;
- $\mathcal{M} \models V_i \rightarrow \mathbf{B}_i V_i$;
- $\mathcal{M} \models \neg V_i \rightarrow \mathbf{B}_i \neg V_i$
- $\mathcal{M} \models [i : a](\mathbf{DONE}_i(a) \wedge [i : b] \neg \mathbf{DONE}_i(a))$
- $\mathcal{M} \models \langle i : a \rangle^{-1} \phi \rightarrow \bigwedge_{a \neq b} [i : b]^{-1} \bot$
- $\mathcal{M} \models \bigvee_{i \in Agt} \langle i : a \cup i : \overline{a} \rangle^{-1} \phi \rightarrow \bigwedge_{i \in Agt} [i : a \cup i : \overline{a}]^{-1} \phi$

The first three entries are standard for $KD45$ (or weak $S5$) models of beliefs [6], forbidding logical inconsistency and allowing positive and negative introspection. The fourth and fifth add positive and negative introspection for Significant Others. It makes sense to claim that if an agent has some Significant Other x if and only if he also believes to have x as Significant Other. The next one shows how action execution is witnessed by the modality **DONE**. The last two validities correspond to linear past. We insert the proof of the last validity, the others are standard or easy to work out.

Proof Assume $\mathcal{M} \models \bigvee_{i \in Agt} \langle i : a \cup i : \overline{a} \rangle^{-1} \phi$. Take an arbitrary pair model-world (M, w). This means that there is an agent j and a pair model-world M', w' such that $(M, w) [\![j : a \cup j : \overline{a}]\!]^{-1} (M', w')$. We know that there is unique (M'', w'') s.t. $(M'', w'') < (M, v)$. Recall that $(M'', w'') < (M, v)$ if there is an event ξ such that $(M'', w'') [\![\xi]\!] (M, v)$. Take ξ to be $j : b$. We know that $M', w' \models \phi$. Uniqueness means that for all other actions c and agents k, $(M, w) [\![k : c]\!]^{-1} (M', w')$. So this is enough to conclude $\mathcal{M} \models \bigwedge_{i \in Agt} [i : a \cup i : \overline{a}]^{-1} \phi$.

4.3 Changing Significant Others

In [36] non standard actions such as those that induce mind changing are described. In the same fashion we would like to describe those actions that update the authority relations among agents. In particular agents should be able to resolve their cognitive dissonance by eliminating Significant Others or welcoming new ones.

We describe the transition function $[\![ ]\!]_R$ for actions $welcome, eliminate, replace$ leaving the treatment of the capability function c [35] that tells us when agents have the internal ability to perform these actions, to future work. These are special actions that transform the models in an intuitive way. The first updates the set of relevant others by adding a new agent. Violation states are updated as specified. The second deletes an agent from such set. The third operation first deletes some agents and after adds new ones to the set.

Definition 1 For some E-KAROUS model

$M = \langle Agt, W, Act, \sigma, \{B_i | i \in Agt\}, \{D_i | i \in Agt\}, Aut, Ag \rangle$ with $w \in W$ and $p, q \in \Pi_0$ be given. We define:

All $\langle M', w'\rangle \in [\![i : \textbf{welcome}(j)]\!]_R \langle M, w\rangle$ are such that:
$M' = \langle Agt, W, Act, \sigma', \{B_i | i \in Agt\}, \{D_i | i \in Agt\}, Aut', Ag\rangle$ with

- $Aut'(i', j', \langle M', w'\rangle) = Aut(i', j', \langle M, w\rangle)$ if $i \neq i'$ or $j \neq j'$ or $w' \neq w$;
- $Aut'(i, j, \langle M', w'\rangle) = \{w | M, w \models V_j\}$;
- $\sigma'(q) = \sigma(q)$ for $q \neq V_j$

After welcoming an agent the authority relation referred to him is updated with the reason of such welcome, while the authority relation concerning the other agents does not change. A corresponding change happens at the propositional level, in such a way that the violation states of the welcomed agent become violation for the agent that performs the welcoming action.

Definition 2 $\langle M', w'\rangle \in [\![i : \textbf{eliminate}(j)]\!]_R \langle M, w\rangle$ are such that:
$M' = \langle Agt, W, Act, \sigma', \{B_i | i \in Agt\}, \{D_i | i \in Agt\}, Aut', Ag\rangle$ with

- $Aut'(i', j', \langle M', w'\rangle) = Aut(i', j', \langle M, w\rangle)$ if $i \neq i'$ or $j \neq j'$ or $w' \neq w$;
- $Aut'(i, j, \langle M', w'\rangle) = \emptyset$;
- $\sigma'(q) = \sigma(q)$ for $q \neq V_j$

Eliminating behaves dually. The agent that is eliminated does not represent an authority anymore, and the change is mirrored at the propositional level. Notice that eliminating an agent is not enough to turn a bad situation in a good one. There may be $n-1$ more eliminating actions needed if we assume $|Agt| = n$.

Definition 3

$$\left[\!\left[i : \textbf{replace}(j, k)\right]\!\right]_R \langle M, w\rangle = [\![(i : \textbf{eliminate}(j); \textbf{welcome}(k)]\!]_R \langle M, w\rangle$$

Replacing is a mere composition of eliminating and welcoming.

Proposition 1 *For any* $M \in \mathcal{M}, q \in \Pi_0, i \neq j, j \neq k, k \neq i$:

- $M \models \textbf{\textit{Sig}}_{i,k} \leftrightarrow [i : \textbf{\textit{welcome}}(j)]\textbf{\textit{Sig}}_{i,k}$;
- $M \models [i : \textbf{\textit{welcome}}(j)]\textbf{\textit{Sig}}_{i,j}$;
- $M \models [i : \textbf{\textit{welcome}}(j)](V_j \rightarrow V_i)$;
- $M \models q \rightarrow [i : \textbf{\textit{welcome}}(j)]q$;
- $M \models \textbf{\textit{Sig}}_{i,k} \leftrightarrow [i : \textbf{\textit{eliminate}}(j)]\textbf{\textit{Sig}}_{i,k}$;
- $M \models [i : \textbf{\textit{eliminate}}(j)]\neg\textbf{\textit{Sig}}_{i,j}$
- $M \not\models [i : \textbf{\textit{eliminate}}(j)](V_j \rightarrow V_i)$;
- $M \models q \rightarrow [i : \textbf{\textit{eliminate}}(j)]q$;
- $M \models \textbf{\textit{Sig}}_{i,j} \rightarrow [i : \textbf{\textit{replace}}(j, k)]\textbf{\textit{Sig}}_{i,k}$;
- $M \models [i : \textbf{\textit{replace}}(j, k)]\neg\textbf{\textit{Sig}}_{i,j}$
- $M \models q \rightarrow [i : \textbf{\textit{replace}}(j, k)]q$

The first four items deal with the welcoming operation. The first of them says that welcoming a new agent does not affect the perception of others; the second simply that welcoming causes an agent to be a Significant Other; the third that the reason of welcome is automatically a violation for the agent; the fourth that the evaluation of the other propositions do not change. The eliminating operations behaves dually, while replacing can be obtained by composing the other two. These validities clearly correspond to the relational changes that define these operations.

5 The dynamics of guilt and shame

5.1 Sense of guilt

Preconditions An agent feels guilty when it observes that what he actively caused was violation for some of his Significant Others.

This means that the agent believes the actual world state he actively chose (that he could avoid) satisfies a violation condition for some agent j which happens to be a Significant Other.

For simplicity, we limit the treatment to atomic cases. It is nevertheless possible to address complex actions. Taken $a \in Act$,

$$\mathbf{B}_i(\mathbf{Sig}_{i,j} \wedge V_j \wedge \mathbf{DONE}_i(a) \wedge \mathbf{P}(\langle i : \overline{a}\rangle \neg V_j)) \rightarrow guilty(i, a, j)$$

For instance if the President believes that the American people are a Significant Other and that he did an action he should not have done and he believes that there was a good alternative that he did not carry out, then he feels guilty.

Reactions In the KARO framework a classical deliberation cycle is assumed [19]. In our case deliberation is a just a special action that updates beliefs, desires, commitments and status of Significant Others by means of the above defined revision actions. For our purpose it is sufficient to know that we can have an action with special effects, as for instance those that welcomed or eliminated Significant Others. Reactions to sense of guilt are influenced by the level of self esteem agents have. We can distinguish two categories, Agt_h, high self esteem agents, which will react providing justifications to their actions and in extreme cases changing the significance they attribute to people. On the other side, Agt_l, low self esteem people will try to find excuses for their actions and to generate reparation goals, that is to perform an action in such a way to avoid further violations.

$$H(i) \wedge \mathbf{B}_i(V_j) \wedge guilty(i, a, j) \rightarrow [deliberate_i](\langle i : \mathbf{eliminate}(j)\rangle(p \vee \neg p) \vee \mathbf{B}_i(\neg V_j)$$

So either i will update authority relations by cancelling j, or he will believe the present state is not violation for j.

On the other hand the following may be said for LSE agents.

$$L(i) \wedge \mathbf{B}_i(V_j) \wedge guilty(i, a, j) \rightarrow [deliberate_i]\mathbf{B}_i(\mathbf{P}[i : \overline{a}]V_j))$$

The low self esteem agent will find excuses for his wrongdoing, for instance he will generate the belief that what it did was unavoidable.

5.2 Shame

Preconditions Shame is the believed lack of a relevant feature, that is the believed incapacity to achieve a value that is important for the agent or for its Significant Others. If only the first is present we will talk of shame towards the self. We are going to formalize shame by considering an agent that believes it is possible to get over a violation state but that there is no capability for him/her to do so.

$$\mathbf{B}_i(\mathbf{Sig}_{i,j} \wedge V_j \wedge \mathbf{DONE}_i(a) \wedge \mathbf{P}([i : \overline{a}]V_j)) \rightarrow shame(i, a, j)$$

So avoiding V_j is something which i is not able to comply. For instance if Bill believes that the American people are a Significant Other and that he did an action he should not have done and he believes that there was no good alternative, then Bill feels ashamed.

Reactions What do ashamed agents do? We distinguish LSE reactions and HSE reactions. Similarly with sense of guilt, the first types of reactions will tend to manipulate the belief base in such a way to remove the belief of incapacity.

$$shame(i, a, j) \wedge i \in L(i) \rightarrow [deliberate_i](\mathbf{B}_i(\neg[i : \overline{a}]V_j)$$

The second type of reactions will try to update authority relations, so that they do not perceive their incapacity as wrong. This is a typical pride reaction.

$$shame(i, a, j) \wedge i \in H(i) \rightarrow [deliberate_i]\langle i : \mathbf{eliminate}(j)\rangle(p \vee \neg p)$$

5.3 Further properties

Our language is powerful enough to distinguish shame and sense of guilt.

$$\mathcal{M} \not\models shame(i, a, j) \leftrightarrow guilty(i, a, j)$$

We are also able to prove that the two feelings are locally incompatible, that is either an agent believes to be a dull knife or a sharp knife.

$$\mathcal{M} \models \neg(shame(i, a, j) \wedge guilty(i, a, j))$$

This is of course to be related with the property of consistent beliefs. If we allow for epistemic accessibility relation to be locally empty, we may conceive situations M, w in which agents believe to be guilty and ashamed.

$$M, w \models B_i\bot \rightarrow shame(i, a, j) \wedge guilty(i, a, j)$$

It has to be noticed that $shame(i, b, j) \wedge guilty(i, a, j)$ is a satisfiable formula, as well as $shame(i, a, k) \wedge guilty(i, a, j)$, even when forcing beliefs to be consistent.

Example The properties of our logic seem to suggest that Bill could not feel both ashamed and guilty for the very same thing. It is then puzzling to read that he felt responsible in its days of shame. We think that indeed Bill must have felt guilty while admitting his responsibility. There was something he could have done and it could have avoid the damage.

$$M, w \models B_{Bill}(\langle kissmonica\rangle^{-1}\langle\overline{kissmonica}\rangle OK)$$

And he and presumably his wife considered what he did wrong.

$$M, w \models V_{Hilary} \wedge \mathbf{Sig}_{Hilary,Bill}$$

Notice that all this amounts to saying

$$M, w \models \mathbf{B}_{Bill}(\mathbf{Sig}_{Hilary,Bill} \wedge V_{Hilary} \wedge \mathbf{DONE}_{Bill}(kissmonica) \wedge \mathbf{P} \times(\langle Bill : \overline{kissmonica}\rangle\neg V_j))$$

And by Propositional Reasoning:

$$M, w \models guilty(Bill, kissmonica, Hilary)$$

However, he explicitly declared both embarrassment and perception of responsibility. The real case is of course more complex than the logical abstractions, nevertheless a model is possible. In fact the problem of Bill was his being both a man (with his personal life and affairs) and a President (with his public life and affairs).

What happened at the presidential level can be described as follows:

$$M, w \models B_{President}(\langle president : lie\rangle^{-1}\langle\overline{president : lie}\rangle OK)$$

We can assume that moral integrity is a value for an American President:

$$\mathcal{M} \models [president : lie][president : a \cup president : \overline{a}]^* V_{Americans}$$

If he lies once he becomes and remains a bad President. We can imagine for instance that he will lose his good reputation. Anyway it is an act he cannot repair: after the lie action there is no action to go back.

This allows us to conclude:

$$M, w \models guilty(President, president : lie, Americans) \wedge$$
$$[president : lie]ashamed(President, president : lie, Americans)$$

Of course a good strategy would have been not to lie. Moreover we suspect

$$M, w \models guilty(President, kissmonica, Americans) \wedge$$
$$[kissmonica]ashamed(President, kissmonica, Americans)$$

to hold, as well. For the Public Opinion it is a good thing to have a President that also in his private life does not cheat.

This in fact might have caused President Clinton to carry out the following LSE action:

$$[deliberate_{President}][president : a \cup president : \overline{a}]V_{Americans}$$

Instead of the HSE reaction:

$$[deliberate_{President}]\neg V_{Americans}$$

On the one hand he felt ashamed towards the American People (that he was not able to eliminate as Significant Other), for a feature that he did not have (moral integrity as a President). On the other he felt responsible towards his family for an action he deliberately committed.

5.4 Collective shame and sense of guilt

In many social interactions, as in the Lewinski case, responsibility for a damage is often shared. So far we could not express the fact that two agent could only work together by avoiding a damage. We can generalize the treatment of emotions by introducing a parallel action in the language.

The semantics of collective action is treated in the appendix, the syntax of Evt is extended in the following way:

$$\xi ::= i : \alpha | \xi \cup \xi | \xi ; \xi | \xi^* | \xi \cap \xi$$

Moreover we introduce new atoms that talk about collective emotions.

For instance $guilty(\{i, j\}, i : a \cap j : b, k)$ means that the set of agents $\{i, j\}$ feels collectively guilty with respect to the event $i : a \cap j : b$, towards the common Significant Other k.[9]

[9] Also here generalizations are possible, think of two agents that feel guilty with respect to two different Significant Others for the very same fact.

To interpret this atom we need to further extend the valuation function in the obvious way: we skip the details for clarity.

In our models the following propositions hold:

$$\mathcal{M} \not\models guilty(\{i, j\}, i : a \cap j : b, k) \leftrightarrow guilty(i, a, k) \wedge guilty(j, b, k)$$

It is satisfiable that a coalition is feels collectively guilty but its components do not.

$$\mathcal{M} \models ashamed(\{i, j\}, i : a \cap j : b, k) \rightarrow ashamed(i, a, k) \wedge ashamed(j, b, k)$$

If a coalition is ashamed then its components are.

$$\mathcal{M} \models guilty(i, a, k) \wedge guilty(j, b, k) \rightarrow guilty(\{i, j\}, i : a \cap j : b, k)$$

If the components of a coalition feel guilty then also the coalition does.

$$M \models (guilty(\{i, j\}, i : a \cap j : b, k) \leftrightarrow guilty(i, a, k) \wedge guilty(j, b, k)) \Leftrightarrow$$
$$M \models ([i : a \cap j : b]\phi \leftrightarrow [i : a]\phi \wedge [i : b]\phi) \text{ is a validity .}$$

In words, individuals are guilty as the coalition they form iff they cannot do more together than what they already could do separately.

We include the proof of the last proposition, the others follow the same pattern:

Proof ($\Leftarrow$)

Assume that $([i : a \cap j : b]\phi \leftrightarrow [i : a]\phi \wedge [i : b]\phi)$ is a validity . This means that taken an arbitrary model M, $M \models [i : a \cap j : b]\phi \leftrightarrow [i : a]\phi \wedge [i : b]\phi$. This in turn means that for any world w the set of successors v via the $[i : a \cap j : b]$ action is equivalent to the set of successors v' via the independent actions $[i : a]$ and $[i : b]$. Suppose now $M, w \models (guilty(\{i, j\}, i : a \cap j : b, k)$. This means that there is an action in $i : \overline{a} \cap j : \overline{b}$ that has $\neg V_k$ as a consequence. By the assumed equivalence this means that $[i : \overline{a}]\neg V_k \wedge [j : \overline{b}]\neg V_k$. We are allowed to conclude that $M, w \models guilty(i, a, k) \wedge guilty(j, b, k))$. $(guilty(\{i, j\}, i : a \cap j : b, k)$ follows by the previous proposition.

($\Rightarrow$)

Assume that $M \models (guilty(\{i, j\}, i : a \cap j : b, k) \leftrightarrow guilty(i, a, k) \wedge guilty(j, b, k))$. So the worlds for which i, j feel guilty together is the same of those for which they feel guilty independently. This is a validity in an arbitrary model, so it is independent of the atomic valuations. So the truth of the formulas are only dependent on the transitions. This is enough to conclude that also the worlds they can reach together and those they can reach independently constitute the same set.

6 Conclusion and future work

In this paper we provided a formal language to describe sense of guilt and shame as social emotions. In order to do this we grounded our work on the cognitive theory of Castelfranchi and Miceli, the psychological theory of Significant Others by Higgins, the rational action theory in the KARO framework by Meyer and colleagues. The cognitive science perspective has allowed us to build an anatomy of these emotions in terms of basic cognitive ingredients such as Beliefs, Goals and Values. We described formally the operations that allow agents to change their evaluations together with the people they take as references, and we connected these to shame, sense of guilt and their dynamics in a Multi Agent System where agents can coordinate and act together.

Much work still needs to be done. Apart from what was already pointed out throughout the paper, we would like to: investigate further the theory of cognitive dissonance and to give a formal characterization of the role of emotions in its resolution; to shed more light on the characterization of emotions by studying the logical models that we used to talk about them: could we rewrite the conditions that trigger these emotions without recurring to past reasoning, but only as in [19] reasoning about the resulting conditions after an action execution? We would like to understand the connection of feeling ashamed and guilty with other feelings like happiness and sadness already formally described in [19] and the agent types defensive and offensive strategies in [22] and [20]. Finally, the Multi Agent view on emotional agents can be profitably connected to already existing game-theoretical frameworks (such as [4]) and in general to Normative Multi Agent Systems [1], that study how norms are enforced and selected in agent societies.

Acknowledgements A lot of thanks to Maria Miceli for the inspiring discussions over norms and emotions. We further thank the anonymous reviewers for their suggestions, especially one of them for starting our exciting case study.

Appendix: Multi Agent action semantics

As in [18] we use synchronicity sets to model concurrency. There synchronicity sets represent those atomic actions that are executed in parallel by the agent. We borrow this idea to model concurrency in a Multi Agent System, arguing in a game-theoretic fashion that the *interaction results from a parallel execution of strategies done by agents*.

Notice that we do not consider groups of agents (or coalitions), as done for instance in Cooperative Game Theory [25]. Further extensions in this direction will be considered.

In extending synchronicity sets for a MAS we label them with the corresponding set of actors, henceforth "labeled synchronicity sets".

Definition 4 A labeled synchronicity set is a sequence $s = \langle a_1, \ldots, a_n \rangle$ that associates to each agent $i (1 \leq i \leq n)$ an action from the action set Act

A labeled synchronicity set can be written in this form:

$$\begin{bmatrix} 1 : a \\ 2 : b \\ \ldots \\ n : c \end{bmatrix}$$

where $n = |Agt|$.

We may abbreviate the sequence into $[\overrightarrow{i : a}]$, for i agent and a action. We denote labeled synchronicity sets with $s, s_1, s_2, \ldots$. We denote the action of an agent i in a labeled synchronicity set s with $s^i = a$ in case $i : a$ is the i-th element of s.

The set of labeled synchronicity sets is $\mathcal{S} = \{[\overrightarrow{i : a}] | a \in Act \text{ and } i \in Agt\}$. $\mathcal{S}$ is finite, provided Agt and Act are.

Sets can be composed, giving rise to traces: $\mathcal{TR} = \{\langle s_1, \ldots, s_n, \ldots \rangle | s_i \in \mathcal{S}\}$. Intuitively traces represent the evolution of the System given a sequence of Multi Agent actions. Traces can be either finite or infinite.

Definition 5 As a semantic counterpart of ; we take the operation $\circ$ that composes traces: $t_1 = \langle s_1, \ldots, s_n \rangle \circ t_2 = \langle s'_m \ldots s'_{m+k} \rangle$ becomes $\langle s_1, \ldots, s_n, s'_m, \ldots, s'_{m+k} \rangle$.

Intuitively, if we do not allow for indeterministic actions, a sequence of synchronicity sets gives rise to a unique trace. Nevertheless we may not know the whole synchronicity set. If we want to speculate on the possible consequences of sequences of actions of an agent or of a subset of agents we need to take into account *the possible answers of the opponents*, that is why it is useful to consider sets of traces. This is also true if we simply have a finite trace at our disposal: the future is open.

We denote with TR, TR_1, TR_2 sets of traces. They can be composed as follows:

$$TR_1 \circ TR_2 = \{tr_1 \circ tr_2 | tr_1 \in TR_1 \text{ and } tr_2 \in TR_2\}.$$

If tr_1 is infinite $tr_1 \circ tr_2 = tr_1$, for $tr_1, tr_2 \in \mathcal{TR}$.

In order to relate traces originated by action execution to transitions we will contruct as in [18] a semantic function $[\![]\!] : Agt \rightarrow Act \rightarrow 2^{\mathcal{TR}}$. To ease notation we will say that the event $x : f$ (i.e. the action f made by x) gets associated to a set of traces $\mathcal{K}$ by the function $[\![]\!]$, formally $\mathcal{K} = [\![x : f]\!]$.

Definition 6 For N the natural numbers, the behaviour of $[\![]\!]$ is described as follows:

- $[\![i : a]\!] = \{s \in \mathcal{S} | s^i = a\}$;
- $[\![i : \overline{a}]\!] = \{s \in \mathcal{S} | s^i \neq a\}$;
- $[\![i : \alpha; \beta]\!] = [\![i : \alpha]\!] \circ [\![i : \beta]\!]$;
- $\left[\!\left[i : \alpha^*\right]\!\right] = \bigcup_{n \in N} \left[\!\left[i : \alpha^n\right]\!\right]$;
- $[\![\xi; \theta]\!] = [\![\xi]\!] \circ [\![\theta]\!]$;
- $[\![\xi \cup \theta]\!] = [\![\xi]\!] \bigcup [\![\theta]\!]$;
- $\left[\!\left[\xi^*\right]\!\right] = \bigcup_{n \in N} \left[\!\left[\xi^n\right]\!\right]$

What is left to do now is to associate action traces with a transition system, that is use the full power of the action expression in the Kripke Models.

Definition 7 Taken $\rho : \mathcal{S} \rightarrow \mathcal{M} \times W \rightarrow 2^{\mathcal{M} \times W}$, where $\mathcal{M}$ is a set of models and W of possible worlds,[10] the behaviour of the function $R : 2^{\mathcal{TR}} \rightarrow \mathcal{M} \times W \rightarrow 2^{\mathcal{M} \times W}$, for $tr_1, tr_2 \in \mathcal{TR}$ is described as follows:

- $R(s)\langle M, w \rangle = \rho(s)\langle M, w \rangle$;
- $R(tr_1 \circ tr_2)\langle M, w \rangle = R(tr_2)(R(tr_1)\langle M, w \rangle)$.

$$R(TR)\langle M, w \rangle = \{\langle M', w' \rangle | \langle M', w' \rangle \in R(tr_1)\langle M, w \rangle \text{ for } tr_1 \in TR\}.$$

We redefine the function $[\![]\!]_R$ as the composition of the function $[\![]\!]$ with the function R.

Definition 8

$$[\![\xi]\!]_R \langle M, w \rangle = R([\![\xi]\!])\langle M, w \rangle.$$

As is clear from the definition we allow for indeterministic actions, that is actions of which terminating executions are not necessarily ending up in a unique state [17].

For convenience, we sometimes view $[\![]\!]_R$ as a functional relation, and we write $\langle M, w \rangle$ $[\![\xi]\!]_R \langle M', w' \rangle$ to mean that from the situation $\langle M, w \rangle$ there is a terminating execution of event ξ that changes the world state into the situation $\langle M', w' \rangle$.

[10] In [18] the performance of actions in a world leads to a collection of worlds, while in [19] it leads to a pair model-world. In this paper we keep both indeterminism of action and model update, just for completeness of representation.

References

1. Agotnes, T., van der Hoek, W., & Wooldridge, M. (2008). Robust normative systems AAMAS '08 In *Proceedings of the 7th international joint conference on Autonomous agents and Multi Agent Systems* (pp. 747–754). Estoril, Portugal.
2. Alur, R., Henzinger, T., & Kupferman, O. (2002). Alternating-time temporal logic. *Journal of the ACM, 49*, 672–713. Preliminary versions appeared in the proceedings of the 38th annual symposium on foundations of computer science (FOCS) (pp. 100–109), IEEE Computer Society Press, 1997.
3. Baltag, A., Moss, L., & Solecki, S. (1999). *The Logic of Public Announcements, Common Knowledge and Private Suspicions*, Technical Report, CWI, VU, Amsterdam.
4. Battigalli, P., & Dufwenberg, M. (2007). Guilt in Games, American Economic Review. *American Economic Association, 97*(2), 170–176.
5. BBC news (1998, September 21 Monday). http://news.bbc.co.uk/2/hi/events/clinton-under-fire/latest-news/176096.stm.
6. Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal logic*. Cambridge tracts in theoretical computer science. Cambridge: Cambridge University Press.
7. Castelfranchi, C. (2004). *Cognitive anatomy of shame and guilt: Differences, functions, defensive Moves*, Talk at EABCT, Manchester.
8. Castelfranchi, C., & Poggi, I. (1990). Blushing as a discourse: Was Darwin wrong? In *Shyness and embarrassment: Perspectives from social psychology* (pp. 230–251). Cambridge, England: Cambridge University Press.
9. CNN Report. (1998). Investigating the President: The trial, http://www.cnn.com/ALLPOLITICS/resources/1998/lewinsky/.
10. Conte, R., & Paolucci, M. (2004). Responsibility for societies of agents, *JASSS*, http://jasss.soc.surrey.ac.uk/7/4/3.html.
11. Damasio, A. (1999). *Descartes' Error: Emotion, Reason and the Human Brain*. New York: G.P. Putnam's Sons.
12. Dastani, M., & Meyer, J.-J. Ch. (2006). Programming agents with emotions. In *17th European conference on artificial intelligence (ECAI 2006)* (pp. 215–219). Amsterdam: IOS Press.
13. Dastani, M., van Riemsdijk, B., Dignum, F., & Meyer, J.-J.Ch. (2004). *A Programming Language for Cognitive Agents Goal Directed 3APL*, Programming Multi Agent Systems, First International Workshop, ROMAS 2003, Melbourne, Australia, July 15, 2003, Selected Revised and Invited Papers. Lecture Notes in Computer Science.
14. de Sousa, R. (2003) *Emotion*, Stanford encyclopedia of philosophy. http://plato.stanford.edu/entries/.
15. Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
16. Grossi, D., Royakkers, L. M. M., & Dignum, F. (2007). Organizational structure and responsibility: An analysis in a dynamic logic of organized collective agency. *Artificial Intelligence and Law, 15*(3), 223–249.
17. Harel, D. (1984). Dynamic Logic. In D. Gabbay & F. Guenther (Eds.), *Handbook of Philosophical Logic Volume II—Extensions of Classical Logic* (pp. 497–604). Dordrecht, The Netherlands: D. Reidel Publishing Company.
18. Meyer, J.-J. Ch. (1988). A different approach to Deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic, 29*(1), 109–136.
19. Meyer, J.-J. Ch. (2004). Reasoning about emotional agents. In R. Lopez de Mantaras & L. Saitta (Eds.), *Proceedings of the 16th European conference on artificial intelligence (ECAI 2004)* (pp. 129–133). Amsterdam: IOS Press.
20. Miceli, M. (1992). How to make someone feel guilty: Strategies for guilt inducement and their goals. *Journal for the Theory of Social Behaviour, 22*, 81–104.
21. Miceli, M., & Castelfranchi, C. (1989). A cognitive approach to values. *Journal for the Theory of Social Behaviour, 19*, 169–194.
22. Miceli, M., & Castelfranchi, C. (1998). How to silence one's conscience: Cognitive defenses against the feeling of guilt. *Journal for the Theory of Social Behaviour, 28*, 287–318.
23. Oatley, K., & Jenkins, J. M. (1996). *Understanding emotions*. Oxford: Blackwell.
24. Ortony, A., Clore, G. L., & Collins, A. (1998). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

25. Pauly, M. (2001). A logic for social software (PhD thesis, ILLC dissertation series, Amsterdam).
26. Pitt, J. (2004). Digital Blush: Towards shame and embarrassment in Multi Agent information trading applications. *Cognition, Technology and Work, 6*, 23–36.
27. Royakkers, L. M. M. (1998). Extending deontic logic for the formalization of legal rules (PhD thesis, Dordrecht, Kluwer Academic Publishers).
28. Shoham, Y. (1990). *Agent-oriented programming*. Stanford, CA 94305: Computer Science Department, Stanford University. (Technical Report STAN-CS-1335-90).
29. Sloman, A. (1990). Motives, mechanisms and emotions. In M. Boden (Ed.), *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press. (pp. 231–247).
30. Staller A., & Petta, P. (2001). Introducing emotions into the computational study of social norms: A first evaluation. *Journal of Artificial Societies and Social Simulation, 4*(1), http://www.soc.surrey.ac.uk/JASSS/4/1/2.html.
31. Tory Higgins, E. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*, 319–340.
32. Turrini, P., Meyer, J.-J. Ch., & Castelfranchi, C. (2007). Rational Agents that Blush. In *Proceedings of the 1st international conference of affective computing and intelligent interaction (ACII 2007)* (pp. 314–325). Berlin Heidelberg: Springer Verlag.
33. Turrini, P., Paolucci, M., & Conte, R. (2006). Social responsibility among deliberative agents. In P. Peppas & A. Perini (Eds.), *Proceedings of Stairs 2006, Loris Penserini* (pp. 38–47). Amsterdam: IOS Press.
34. van Benthem, J. (2005). Where is logic going, and should it? In E. Bencivenga (Ed.), *What is to be Done in Philosophy*.
35. van der Hoek, W., van Linder, B., & Meyer, J.-J. Ch. (1994). A logic of capabilities. *Lecture Notes in Computer Science, 813*, 366–413.
36. van Linder, B., van der Hoek, W., & Meyer, J.-J. Ch. (1995). Actions that make you change your mind. *Knowledge and Belief in Philosophy and Artificial Intelligence*, pp. 103–146.
37. van Linder, B., van der Hoek, W., & Meyer, J.-J. Ch. (1998). An integrated modal approach to rational agents. In M. Wooldridge & A. Rao (Eds.), *Foundations of Rational Agency, Applied Logic Series 14*. Dordrecht: Kluwer.