Emotions as Heuristics for Rational Agents

Bas R. Steunebrink

Mehdi Dastani

John-Jules Ch. Meyer

Department of Information and Computing Sciences, Utrecht University

Technical Report UU-CS-2007-006

www.cs.uu.nl
ISSN: 0924-3275

Emotions as Heuristics for Rational Agents

Bas R. Steunebrink
Department of Information and Computing Sciences
Utrecht University
Utrecht, The Netherlands
bass@cs.uu.nl

Mehdi Dastani
Department of Information and Computing Sciences
Utrecht University
Utrecht, The Netherlands
mehdi@cs.uu.nl

John-Jules Ch. Meyer
Department of Information and Computing Sciences
Utrecht University
Utrecht, The Netherlands
jj@cs.uu.nl

ABSTRACT

In this paper we argue that a model of human emotions provide reasonable and useful heuristics for 1) reducing and controlling nondeterminism involved in an agent's decision making process, 2) flexible cooperation and coordination between agents, and 3) building efficacious human-agent interfaces to facilitate the interaction between human users and artificial agents in hybrid multi-agent systems. In order to incorporate heuristics that are inspired by human emotions in an agent model, we first discuss an existing model of human emotions proposed by Ortony, Clore & Collins ("OCC") and present a formalization of that model. This is done by introducing a logical language and its semantics that are used to specify an agent model in terms of mental attitudes including emotions. The emotions in this model function as heuristics as they constrain an agent's model. We show in this model how emotions help reducing nondeterminism in an agent's decision making process.

1. INTRODUCTION

Many branches of computer science in general and multi-agent systems in particular involve problems that need heuristics to manage their complexities. We will discuss three problems in multi-agent systems for which we propose heuristics inspired by human emotions. These problems are related to the autonomy of individual agents, the cooperation and coordination between agents, and the believability of and efficacious interaction with agents in hybrid multi-agent systems.

In multi-agent systems, individual agents are assumed to be *autonomous*. This implies that they should have the capability to deliberate and decide which actions or plans to perform in order to reach their objectives. For many practical applications and for building interpreters for agent programming languages, standard decision-theoretic concepts (e.g., probability and utility) and rules (e.g., maximizing expected utility) will leave an agent with multiple actions or plans with equal or near-equal preference. This gives rise to nondeterminism in an agent's decision making, which is a problem if the agent has to choose only one of the options at each moment in time (e.g., the interpreter of a BDI agent needs to decide which action to select at each deliberation cycle [3]). Moreover, individual agents need to *cooperate* and *coordinate* their actions in order to achieve their global (social) objectives, often called the objectives of a multi-agent system. Interaction protocols and mechanisms are the common means to specify the cooperation and coordination of individual agents to guarantee the achievement of their global objectives. However, the use of interaction protocols imposes restrictive constraints on the behavior of individual agents and thereby limits their autonomy. On the other hand, the absence of interaction protocols brings the problem of nondeterminism back to individual agents, because there may be many options for interaction with other agents. It should be noted that an interaction protocol can in fact be considered as a way to solve the problem of nondeterminism. We believe that for many (non-critical) applications of multi-agent systems the restrictive interaction protocols can be replaced by more flexible and more general policies or heuristics. Finally, in hybrid multi-agent systems, where human users are also considered as constituting agents, efficacious *interfaces* are needed to enhance and optimize the complex interactions between humans and artificial agents. Also, as humans intuitively assign a personality, motives, and an affective state to any system displaying complex behavior, the challenge in human-agent interaction lies with living up to these expectations of the human user and thus creating *believable* agent behaviors.

In this paper, we argue that as long as heuristics are needed for solving the three mentioned problems in multi-agent systems, a model of affect based on human-inspired emotions may provide reasonable and useful heuristics. Moreover, we argue that using a model of affect is especially beneficial because it is a single mechanism that functions as a heuristic for all three of these problems. We propose using heuristics inspired by human emotions to (partially) solve the aforementioned problems in a functional way. Specifically, we propose to formalize the existing functional OCC model of human emotions proposed by Ortony, Clore & Collins [12] and incorporate it in the existing multi-agent KARO framework [8, 9]. This is done by extending the syntax and semantics of the KARO framework with emotions. The functional structure of emotions as proposed in the OCC model are then considered as heuristics that constrain the model of agency suggested by the KARO framework. We show that the resulting constraint model will reduce and control the nondeterminism involved in artificial agents.

The outline of this paper is as follows. We start in section 2 by motivating how a model of human emotions can be used to (partially) solve the three aforementioned problems in multi-agent systems. Furthermore, we introduce the model of human emotions that we are using and discuss relevant related work. In section 3 we introduce the agent specification language (and its semantics) in which we formalize our model of emotions. Part of this formalization is presented in section 4 and interesting properties of the formalized emotions are discussed in section 5. The effects of

emotions on the deliberation of an agent are discussed and formalized in section 6. Finally, section 7 concludes this paper.

2. MOTIVATION AND RELATED WORK

In psychological studies, the emotions that influence the deliberation and practical reasoning of an agent are called secondary (or deliberative) emotions [12, 11]. They are considered as heuristics for preventing excessive deliberation [2]. Meyer & Dastani [8, 3] propose a functional approach to describe the role of secondary emotions in practical reasoning. According to this functional approach, an agent is assumed to execute domain actions in order to reach its goals. The effects of these domain actions cause and/or influence the appraisal of emotions according to a human-inspired model. These emotions in turn influence the deliberation operations of the agent, functioning as heuristics for determining which domain actions have to be chosen next, which completes the circle.

The OCC model describes a hierarchy that classifies 22 secondary emotions. The hierarchy contains three branches, namely emotions concerning aspects of objects (e.g., love and hate), actions of agents (e.g., pride and admiration), and consequences of events (e.g., joy and pity). At the bottom, the second and third branch combine to form a fourth group of compound emotions, namely emotions concerning consequences of events *caused* by actions of agents (e.g., gratitude and anger). Because the objects of all these emotions (i.e. objects, actions, and events) correspond to notions commonly used in agent models (i.e. agents, plans, and goal accomplishments, respectively), this makes the OCC model suitable for use in the deliberation and practical reasoning of artificial agents. It should be emphasized that emotions are not used to describe the entire cognitive state of an agent (as in "the agent is wholly happy"); rather, emotions are always relative to individual objects, actions, and events, so an agent can be joyous about event X and at the same time be distressed about event Y.

The OCC model defines both qualitative and quantitative aspects of emotions. Qualitatively, it defines the conditions that *elicit* each of the emotions; quantitatively, it describes how there is a potential, threshold, and intensity associated with each elicited emotion and what are the variables affecting these quantities. For example, the compound emotion *gratitude* is qualitatively specified as "approving of someone else's praiseworthy action and being pleased about the related desirable event," whereas the variables affecting its (quantitative) intensity are 1) the degree of judged praiseworthiness, 2) the unexpectedness of the event, and 3) the degree to which the event is desirable.

The OCC model has previously been used for emotion synthesis (often partially adapted to the research domain), for example for modeling personalities in social relationships [6, 5], facial expressions for poker playing agents [7], and believable animated characters [1, 10] (cited in Picard [13]).

We propose to use the OCC model for creating heuristics to control the nondeterminism arising in rational / deliberative agents. We will use the OCC model because of its suitability for formalization, but we plan to broaden our research with alternative theories of emotion. An implementation of the OCC model can be achieved by first formalizing the eliciting conditions of these emotions in terms of beliefs, goals, actions, and plans. Then the formal model can be extended with a quantitative model for these emotions, capable of handling emotion potentials, thresholds, and intensities as described by OCC. Finally, the deliberation process of an agent can be extended with heuristics based on the effects of these emotions.

According to Sloman [14], the OCC model can be classified as a *broad* but *shallow* model of emotions. It is broad because it covers a wide range of emotions, but it is shallow because it does not provide a survey of the types of information processing architectures and e.g. the kinds of states, emotions, and experiences they support. However, for our purposes a shallow model of emotions is adequate, because we do not intend to reconstruct a human mind, but rather use emotions as heuristics to develop a directly implementable and usable agent system.

Our use of the OCC model is motivated by the three problems identified in the introduction. As an example of how emotions modeled after the OCC model might reduce nondeterminism in an agent's deliberation and decision making process, suppose an agent generates a plan to achieve a goal. Because the agent is now committed to the plan, it is pleased with the prospect of achieving the associated goal, so it *hopes* to achieve the goal using the plan. However, as soon as the execution of the plan fails (e.g., the execution of some action fails or its effect is not perceived), the agent will also experience *fear* with respect to the plan of not achieving the goal. As soon as the intensity of the fear becomes greater than the intensity of the hope, the agent may drop the plan and start replanning to achieve the goal. Of course, if given multiple revised plans to choose from, the agent will prefer new plans that have previously caused it to experience *relief*, while avoiding those that have previously resulted in the emotion *fears-confirmed*. In this way, the incorporated model of emotions indicates which plans need to be dropped or adopted and thereby helps to reduce the nondeterminism involved in an agent's decision making process.

If agents, which have to cooperate with each other, know how they work internally, they can anticipate expected actions. Note that this assumption is realistic only for closed (not open) multi-agent systems where the design of agents is known. For example, if an agent predicts that a cooperating agent cannot perform its current plan which contributes to the achievement of a common goal and is about to drop it, then the agent may be able to take anticipatory actions in order to prevent the imminent failure of the other agent. We believe that, if all agents were using affective mechanisms as heuristics for their decision making, this kind of cooperation can be achieved in much the same way as humans do in their interactions. Specifically, an agent could accomplish this by mapping the perceived state of another agent to its own affective mechanisms (in our proposal, modeled after OCC) and approximate the affective state of the other agent. It can then predict what actions the other agent is most likely to perform next by running its own affective decision making heuristics. This is what humans do in their interactions; building a model of another person's affective state and predicting how the other person might react. If agents were using the same affective mechanisms, they could do the same type of (implicit) cooperation and coordination.

Finally, humans are experts in social reasoning and behavior. If artificial agents were to use the same affective mechanisms as humans, such as those described by OCC, then humans would have an intuitive and efficacious way of interacting with them. We believe that the use of a human inspired affect model in artificial agents makes the interac-

tion between them and humans more efficacious and natural. A reason is that such agents would be able to live up to the affective expectations of the human to a far greater extent. An affective agent could also present information to its user according to its affective state, further facilitating an intuitive mode of interaction. Consequently, this could greatly increase the *believability* of agents. Conversely, if an agent had affective mechanisms like a human, then it could also map the perceived affective state of a human to its own affective mechanism and choose more suitable and anticipatory actions, just like in multi-agent systems as described above.

It should be noted that previous work on specifying and implementing emotions carried out by Meyer [8] and Dastani [4] follows Oatley & Jenkins' model of emotions [11] and comprises only four emotions: *happy*, *sad*, *angry* and *fearful*. Each emotion functions as a *label* of an aspect of an agent's cognitive state. Similar to our approach, the deliberation of an agent behaves in accordance with the heuristics associated with these four emotions. Dastani & Meyer [4] have defined transition semantics for their emotional model, which we also intend to do for our formalization of the OCC model. However, we intend to formalize the quantitative aspects of emotions as well, which was not considered in the purely logical model of Dastani & Meyer.

3. LANGUAGE AND SEMANTICS

We use KARO [8, 9] as a framework for the formalization of the 22 emotions of the OCC model. The KARO framework is a mixture of dynamic logic, epistemic / doxastic logic, and several additional (modal) operators for dealing with the motivational aspects of artificial agents. We present a modest extension of the KARO framework, so that the eliciting conditions of the emotions of the OCC model can be appropriately modeled. We have completed the formalization of a qualitative model of the 22 emotions in the OCC model, but because of space limitations we cannot present this entire formalization here. Instead, we will focus on *hope* and *fear* alone in this paper. We have picked these two emotions because of their pivotal role in reducing nondeterminism in agent implementations.

The KARO framework is designed to specify goal-directed agents; however, in contrast to KARO, we do not allow arbitrary formulas as (declarative) goals and define a goal as a conjunction of literals, where each literal represents a subgoal. This is because we want agents to be able to break up their goals into subgoals to determine which parts of a goal have already been achieved and which subgoals have yet to be pursued. In order to record the set of achieved subgoals, we will introduce the notion of accomplishment. Furthermore, we require goals to be consistent (individually, not mutually) and non-empty (i.e. there must be at least one literal in the conjunction).

When formalizing the branch (of the OCC hierarchy) of emotions concerning consequences of events, we will translate OCC's notion of an event as the accomplishment or undermining of a goal (or part thereof), because these are the kinds of events telling an agent how its goals and plans toward them are progressing. So such events provide reasonable moments in time at which an agent can decide whether or not its plan(s) need revising. This decision is then influenced by the emotion appraised by the event. For example, replanning may be triggered when fear for failure of the plan to reach the goal is greater than hope for accomplishment of the goal.

Although it is desirable to be able to talk about empty accomplishments, it does not make sense to allow empty goals. Therefore we will define two sets of (consistent) conjunctions below, their difference being the inclusion of the empty conjunction. We will not go into details about how and when goals can be broken up into subgoals, because the emotions presented in this paper (i.e. hope and fear) do not require the notion of accomplishment to be formally defined.

Definition 3.1. (Consistent conjunctions). Let $\mathcal{P}$ be a set of atomic propositions and $Lits = \mathcal{P} \cup \{ \neg p \mid p \in \mathcal{P} \}$ be the set of literals. With respect to the conjunction and disjunction of the empty set, let $\bigwedge \emptyset = \top$ and $\bigvee \emptyset = \bot$, where $\bot$ stands for falsum and $\top$ for verum. Then $\mathcal{K}$ is the set of all consistent conjunctions of literals, and $\mathcal{K}'$ does not contain the empty conjunction:

$$\mathcal{K} = \{ \bigwedge \Phi \mid \Phi \subseteq Lits,\ \Phi \nvdash \bot \} \tag{1}$$

$$\mathcal{K}' = \mathcal{K} \setminus \{\top\} \tag{2}$$

Thus the empty conjunction is denoted by $\top$, and note that $\top \in \mathcal{K}$ whereas $\top \notin \mathcal{K}'$.

When formalizing the branch (of the OCC hierarchy) of emotions concerning actions of agents, we will translate OCC's notion of actions as plans consisting of domain actions and sequential compositions of actions. Other types of actions that are often used in dynamic logic (e.g., tests, conditional statements, loops) are outside the scope of this paper. Note that besides domain actions that can be performed by agents, we also distinguish deliberation operations (e.g., operations for selecting and applying reasoning rules and for selecting plans) as actions that can be performed by agents.

Definition 3.2. (Plans). Let $\mathcal{A}$ be a set of atomic domain actions. The set *Plans* of plans consists of all actions and sequential compositions of actions. It is the smallest set closed under:

- If $\alpha \in \mathcal{A}$ then $\alpha \in Plans$.
- If $\alpha \in \mathcal{A}$ and $\pi \in Plans$ then $(\alpha; \pi) \in Plans$.

When formalizing the branch (of the OCC hierarchy) of emotions concerning objects, we only consider agents as objects, because there are no other notions in our framework that could reasonably be regarded as objects. Emotions concerning agents as objects and actions of other agents can be used to influence the social behavior of an agent; however, these emotions are not treated in this paper and will be a topic of future work.

We define an *emotional fluent* for each of the 22 emotions of the OCC model. The emotions are outlined below such that each row contains two emotions that are defined by OCC to be each other's opposites, with the left column displaying the positive emotions and the right column displaying the negative emotions (for agent i). It should be noted that it is allowed for an agent to have 'mixed feelings,' i.e. it can experience opposing emotions simultaneously. However, our model will ensure that the objects of opposing emotions are distinct (e.g., an agent can experience both joy and distress in response to some event, but the objects of these two emotions will concern different parts of the event).

Definition 3.3. (Emotional fluents). Let $\mathcal{G}$ be a set of agent names. The set *Emotions* is the set of emotional fluents, which is defined as follows:

$$
\begin{aligned}
&\textit{Emotions} = \\
&\{\ \mathbf{joy}_i(\kappa), && \mathbf{distress}_i(\kappa), \\
&\ \ \mathbf{hope}_i(\pi,\kappa), && \mathbf{fear}_i(\pi,\neg\kappa), \\
&\ \ \mathbf{satisfaction}_i(\pi,\kappa), && \mathbf{disappointment}_i(\pi,\kappa), \\
&\ \ \mathbf{relief}_i(\pi,\neg\kappa), && \mathbf{fears\text{-}confirmed}_i(\pi,\neg\kappa), \\
&\ \ \mathbf{happy\text{-}for}_i(j,\kappa), && \mathbf{resentment}_i(j,\kappa), \\
&\ \ \mathbf{gloating}_i(j,\kappa), && \mathbf{pity}_i(j,\kappa), \\
&\ \ \mathbf{pride}_i(\alpha), && \mathbf{shame}_i(\alpha), \\
&\ \ \mathbf{admiration}_i(j,\alpha), && \mathbf{reproach}_i(j,\alpha), \\
&\ \ \mathbf{love}_i(j), && \mathbf{hate}_i(j), \\
&\ \ \mathbf{gratification}_i(\alpha,\kappa), && \mathbf{remorse}_i(\alpha,\kappa), \\
&\ \ \mathbf{gratitude}_i(j,\alpha,\kappa), && \mathbf{anger}_i(j,\alpha,\kappa) \\
&\ \mid i,j \in \mathcal{G},\ i \neq j,\ \alpha \in \mathcal{A},\ \pi \in \textit{Plans},\ \kappa \in \mathcal{K}'\ \}
\end{aligned}
\tag{3}
$$

We now have all ingredients necessary to modify the KARO framework and construct an agent specification language. This language contains operators for belief (**B**), goals (**G**), (cap)ability (**A**), commitment (**Com**), and action (**do**).

Definition 3.4. (Language). Let the sets $\mathcal{P}$, $\mathcal{K}'$, *Plans*, $\mathcal{G}$, and *Emotions* be defined as above. The agent specification language $\mathcal{L}$ is the smallest set closed under:

- If $p \in \mathcal{P}$ then $p \in \mathcal{L}$.
- If $\varphi_1, \varphi_2 \in \mathcal{L}$ then $\neg\varphi_1, (\varphi_1 \wedge \varphi_2) \in \mathcal{L}$.
- If $\varphi \in \mathcal{L}$ and $i \in \mathcal{G}$ then $\mathbf{B}_i\varphi \in \mathcal{L}$.
- If $\kappa \in \mathcal{K}'$ and $i \in \mathcal{G}$ then $\mathbf{G}_i\kappa \in \mathcal{L}$.
- If $\pi \in \textit{Plans}$ and $i \in \mathcal{G}$ then $\mathbf{A}_i\pi, \mathbf{Com}_i(\pi) \in \mathcal{L}$.
- If $\pi \in \textit{Plans}$ and $\varphi \in \mathcal{L}$ and $i \in \mathcal{G}$ then $[\mathbf{do}_i(\pi)]\varphi \in \mathcal{L}$.
- If $\epsilon \in \textit{Emotions}$ then $\epsilon \in \mathcal{L}$.

We also use the propositional connectives $\vee$, $\rightarrow$, and $\leftrightarrow$ with their usual interpretation. $\mathbf{B}_i\varphi$ means agent i believes in φ; $\mathbf{G}_i\kappa$ means agent i has the (declarative) goal to accomplish κ; $\mathbf{A}_i\pi$ means agent i has the ability to perform π; $\mathbf{Com}_i(\pi)$ means agent i is committed to performing π; $[\mathbf{do}_i(\pi)]\varphi$ means φ holds after agent i has performed π. For convenience, subscript agent indices (e.g., i and j) are omitted if the formula in question concerns only a single agent. We use $\langle\,\cdot\,\rangle$ as the dual of $[\,\cdot\,]$ for the **do** operator. We denote the execution of the deliberation operations as $[\mathbf{do}(\textit{deliberate})]$.

With respect to the semantics of $\mathcal{L}$, we model the belief and action operators using Kripke semantics, while using sets for ability, commitment, goals, and emotional fluents. The semantics of actions are defined over the Kripke models of belief, because actions may change the mental state of an agent. We deviate from KARO by defining the goal operator **G** in terms of a given set of goals, instead of an abbreviation of a selected, implementable, but unfulfilled desire (or wish) [9]. This greatly simplifies our language, because it is not necessary to model selection (usually denoted by $\mathbf{C}_i\varphi$) and implementability (usually denoted by $\Diamond_i\varphi$) and their dynamics as done in KARO.

Definition 3.5. (Semantics). Let the sets $\mathcal{P}$, $\mathcal{K}'$, $\mathcal{A}$, *Plans*, and $\mathcal{G}$ be defined as above. The semantics of the belief and action operators are given by Kripke structures of the form $\mathrm{M} = \langle S, \vartheta, R_{\mathbf{B}}\rangle$ and $\langle \Sigma, R_{\mathcal{A}}\rangle$, respectively, where

- S is a non-empty set of states (or worlds);
- $\vartheta : S \rightarrow \wp(\mathcal{P})$ is a truth assignment function per state;
- $R_{\mathbf{B}} : \mathcal{G} \times S \rightarrow \wp(S)$ is an accessability relation on S for the belief modality of an agent. $R_{\mathbf{B}}$ is assumed to be serial, transitive, and euclidean;
- Σ is the set of possible model–state pairs. A model–state pair is denoted as (M, s), where $\mathrm{M} = \langle S, \vartheta, R_{\mathbf{B}}\rangle$ as above and $s \in S$;
- $R_{\mathcal{A}} : \mathcal{G} \times \textit{Plans} \times \Sigma \rightarrow \wp(\Sigma)$ is an accessability relation on Σ, encoding the behavior of actions of an agent. $R_{\mathcal{A}}(i,\pi)$ (for $\pi \in \textit{Plans}$) is defined as usual in dynamic logic by induction from a given base case $R_{\mathcal{A}}(i,\alpha)$ (for $\alpha \in \mathcal{A}$), i.e. $R_{\mathcal{A}}(i,\alpha;\pi) = R_{\mathcal{A}}(i,\alpha) \bullet R_{\mathcal{A}}(i,\pi)$.

The semantics of ability, commitment, goals, and emotions are given by means of structures of type $\langle \mathcal{C}, Ag, \Gamma, E\rangle$, where

- $\mathcal{C} : \mathcal{G} \times \Sigma \rightarrow \wp(\textit{Plans})$ is a function that returns the set of actions that an agent is capable of performing per model–state pair;
- $Ag : \mathcal{G} \times \Sigma \rightarrow \wp(\textit{Plans})$ is a function that returns the set of plans that an agent is committed to (are on an agent's 'agenda') per model–state pair;
- $\Gamma : \mathcal{G} \times \Sigma \rightarrow \wp(\mathcal{K}')$ is a function that returns the set of goals that an agent has per model–state pair;
- $E = \langle \textit{Joy}, \textit{Distress}, \textit{Hope}, \textit{Fear}, \ldots, \textit{Anger}\rangle$ is a structure of 22 functions indicating per model–state pair which emotions are being experienced by an agent.

Note that *Hope* and *Fear* are semantic functions designed to define the semantics of the syntactic emotional fluents **hope** and **fear**. It is *crucial* to note that the functions in E are constrained by the emotion axioms that we define according to the OCC model, i.e. formulas (6) and (7) in this paper. Because we will only be treating hope and fear in this paper, we will only define the semantics, interpretation, and emotion axioms of these two emotions. The emotion functions in E have the following types:

$$
\begin{aligned}
\textit{Hope} &: \mathcal{G} \times \Sigma \rightarrow \wp(\textit{Plans} \times \mathcal{K}') \\
\textit{Fear} &: \mathcal{G} \times \Sigma \rightarrow \wp(\textit{Plans} \times \mathcal{K}^{\neg}) \\
&\ \ \vdots
\end{aligned}
$$

where $\mathcal{K}^{\neg} = \{\,\neg\kappa \mid \kappa \in \mathcal{K}'\,\}$. They have to be defined per agent ($\mathcal{G}$) and model–state pair (Σ); their mappings can be directly derived from Equation (3). The semantics of the other emotions are omitted in this paper, but they are easily reconstructed by analogy. Furthermore, it is assumed that an action/plan π is removed from an agent's agenda Ag as soon as the agent has executed π, which is expressed by the following constraint:

$$
\begin{aligned}
\pi \in Ag(i)(\mathrm{M}, s)\ \&\ (\mathrm{M}', s') \in R_{\mathcal{A}}(i,\pi)(\mathrm{M}, s) \Rightarrow \\
\pi \notin Ag(i)(\mathrm{M}', s')
\end{aligned}
\tag{4}
$$

This constraint can be read as follows: if π is on the agenda Ag of agent i in state s of model M and executing π leads to the new state s' of model M′, then π will not be on the agenda Ag of agent i in state s'. Of course an agent could have put a new instance of plan π on its agenda after performing the 'old' π, but we assume this does not violate the constraint above, because we treat these plans as different *instantiations* of π. Finally, note that we do *not* assume $\Gamma(i)(\mathrm{M}, s) \nvdash \bot$, so goals may be mutually inconsistent.

Having defined the semantic operators, we can present how formulas in $\mathcal{L}$ are interpreted.

Definition 3.6. (Interpretation of formulas). Let $\mathrm{M} = \langle S, \vartheta, R_{\mathbf{B}} \rangle$, $\langle \Sigma, R_{\mathcal{A}} \rangle$, and $\langle \mathcal{C}, Ag, \Gamma, E \rangle$ be structures defined as above. Formulas in language $\mathcal{L}$ are interpreted in model–state pairs as follows:

$$
\begin{aligned}
&\mathrm{M}, s \models p && \Leftrightarrow\ p \in \vartheta(s) \quad \text{for } p \in \mathcal{P} \\
&\mathrm{M}, s \models \neg\varphi && \Leftrightarrow\ \mathrm{M}, s \not\models \varphi \\
&\mathrm{M}, s \models \varphi_1 \wedge \varphi_2 && \Leftrightarrow\ \mathrm{M}, s \models \varphi_1 \ \&\ \mathrm{M}, s \models \varphi_2 \\
&\mathrm{M}, s \models \mathbf{B}_i\varphi && \Leftrightarrow\ \forall s' \in R_{\mathbf{B}}(i)(s) : \mathrm{M}, s' \models \varphi \\
&\mathrm{M}, s \models \mathbf{G}_i\kappa && \Leftrightarrow\ \kappa \in \Gamma(i)(\mathrm{M}, s) \\
&\mathrm{M}, s \models \mathbf{A}_i\pi && \Leftrightarrow\ \pi \in \mathcal{C}(i)(\mathrm{M}, s) \\
&\mathrm{M}, s \models \mathbf{Com}_i(\pi) && \Leftrightarrow\ \pi \in Ag(i)(\mathrm{M}, s) \\
&\mathrm{M}, s \models [\mathbf{do}_i(\pi)]\varphi && \Leftrightarrow \\
& && \qquad \forall(\mathrm{M}', s') \in R_{\mathcal{A}}(i, \pi)(\mathrm{M}, s) : \mathrm{M}', s' \models \varphi \\
&\mathrm{M}, s \models \mathbf{hope}_i(\pi, \kappa) && \Leftrightarrow\ (\pi, \kappa) \in \mathit{Hope}(i)(\mathrm{M}, s) \\
&\mathrm{M}, s \models \mathbf{fear}_i(\pi, \neg\kappa) && \Leftrightarrow\ (\pi, \neg\kappa) \in \mathit{Fear}(i)(\mathrm{M}, s) \\
& && \qquad \vdots
\end{aligned}
$$

Note that we evaluate formulas in state s of model M. The Kripke structure $\langle \Sigma, R_{\mathcal{A}} \rangle$ is then used for the interpretation of $[\mathbf{do}_i(\pi)]\varphi$ formulas. In the rest of this paper, we will express that some formula φ is a validity (i.e. $\forall(\mathrm{M}, s) \in \Sigma : \mathrm{M}, s \models \varphi$) simply as $\models \varphi$.

Finally, we define a notion of possible intention equivalent to the one by Meyer [8]. An agent has the possible intention to perform plan π in order to accomplish κ if and only if it believes that 1) it has the ability to perform π, 2) κ is a goal of the agent, and 3) the execution of π possibly leads to a state where κ holds.

Definition 3.7. (Possible intention). The possible intention I to perform π in order to accomplish κ is defined as:

$$I(\pi, \kappa) \leftrightarrow \mathbf{B}(\mathbf{A}\pi \wedge \mathbf{G}\kappa \wedge \langle \mathbf{do}(\pi) \rangle \kappa) \tag{5}$$

The framework specified above is suitable for formalizing the eliciting conditions of the emotions from the OCC model. We are also developing a quantitative model capable of modeling the intensities, thresholds, and potentials of emotions and their interactions, as described by the OCC model. However, because of space limitations, we cannot present a full quantitative model incorporating all these aspects here. For the example emotions described in this paper (i.e. **hope** and **fear**), we omit the treatment of potentials and thresholds. Furthermore, we restrict ourselves to two intensity values: *low* and *high*. This yields the minimal model required for showing the interplay between hope and fear as described by OCC.

Definition 3.8. (Emotion intensity). The partial function *intensity* assigning intensities to emotions is declared as:

$$\mathit{intensity} : \mathcal{G} \times \Sigma \times \mathit{Emotions} \rightarrow \{\mathit{low}, \mathit{high}\}$$

When supplied with an emotion, this function determines its intensity value. So the *intensity* function has at least 22 definitions (one for each emotion type), of which we will define the **hope** and **fear** cases in section 6. Furthermore, $\mathit{intensity}(i)(\mathrm{M}, s)(\epsilon)$ is undefined if $\mathrm{M}, s \not\models \epsilon$. The *intensity* function is defined per agent and model–state pair; however, for convenience we will hence omit these two arguments.

4. A FORMAL MODAL OF EMOTIONS

The OCC model provides for each emotion (among others) a concise definition in a single sentence and a list of variables affecting the intensity of the emotion in question. Below we will repeat OCC's definitions of *hope* and *fear* (given in [12], page 112) and show how they can be formalized in the language we have just defined. In section 6 we will deal with the intensity part of these emotions.

Hope: According to OCC, *hope is being pleased about the prospect of a desirable event.* From the viewpoint of a goal-directed agent, a *desirable event* can be translated to the accomplishment of a goal or part thereof, whereas the *prospect* can be translated to 'having' a plan for accomplishing that goal. More specifically, we require the agent to intend to perform this plan and to be committed to it. An agent that is *being pleased* about the prospect of a desirable event should act according to this mental state, so here we are hinted at a possible heuristic that can be associated with the emotion hope, namely to keep the intention and commitment while this emotion is strong enough. What exactly it means for hope to be strong enough will be formalized in section 6.

Thus phrased in our language, an agent hopes to achieve some goal using some plan if and only if it intends to perform the plan for the goal and is committed to the plan. The objects of the hope emotion are then the goal that the agent intends to achieve and the plan to which it is committed. We thus arrive at the following definition for hope:

$$\mathbf{hope}(\pi, \kappa) \leftrightarrow (I(\pi, \kappa) \wedge \mathbf{Com}(\pi)) \tag{6}$$

It is important to note the use of the bi-implication, because it allows for the derivation of interesting properties in section 5.

Fear: According to OCC, *fear is being displeased about the prospect of an undesirable event.* However, OCC note that if one experiences hope with respect to the prospect of a desirable event, then the absence of that event will be undesirable to the same degree. In other words, hope and fear are complementary emotions. This means that the intensities associated with hope and fear with respect to the same prospect and event have to sum to a constant. Note that this is different from what we called opposing emotions.

Because we have translated a desirable event as the accomplishment of a goal (or part thereof), an *undesirable event* will constitute the failure to achieve that goal (or part thereof). So fear will arise when the complement of an event hoped for becomes probable (this is the *prospect* part). An agent that is *being displeased* about the prospect of an undesirable event should start considering alternatives in order to ensure that it is the desirable event which will be achieved. Again, how exactly this can be done will be formalized in section 6.

Thus phrased in our language, an agent fears the failure to achieve some goal using some plan if and only if it hopes the plan will achieve the goal but it believes that it may not. The objects of the fear emotion are then the plan from the corresponding hope emotion and the negation of the goal that it is hoping to achieve. We thus arrive at the following definition for fear:

$$\mathbf{fear}(\pi, \neg\kappa) \leftrightarrow (\mathbf{hope}(\pi, \kappa) \wedge \mathbf{B}\langle \mathbf{do}(\pi) \rangle \neg\kappa) \tag{7}$$

Because fear is the complement of hope, hope must be a precondition of fear. The other precondition, namely that the complement of the event hoped for has become probable,

is expressed as the belief that the execution of the intended plan *may* fail to achieve the desired event (note the angled brackets). As with the definition of hope, it is also important to note the use of the bi-implication. Properties that can consequently be derived are explored in Section 5.

It should be emphasized that the two emotion axioms above act as constraints on the functions *Hope* and *Fear* in the semantics of our language. The fact that these two axioms look like mere abbreviations of formulas (as in the case of Equation (5) for possible intention) is coincidental. In our complete qualitative formalization of OCC, most emotional fluents cannot simply be written on the left hand side of a bi-implication.

5. PROPERTIES OF EMOTIONS

Having defined hope and fear in our formal model, we can check whether we can derive interesting properties from these definitions and whether the derivable properties are intuitive. In this section we will discuss several propositions; their proofs can be found in the appendix.

$\models \mathbf{hope}(\pi, \kappa) \rightarrow [\mathbf{do}(\pi)]\neg\mathbf{hope}(\pi, \kappa)$ (8)

Hope only lasts for the duration of the prospect. As soon as the agent has performed plan π with which it hoped to achieve goal κ, the hope disappears, because it is no longer committed to π. This follows almost directly from constraint (4) in combination with definition 3.6, validating $\models \mathbf{Com}(\pi) \rightarrow [\mathbf{do}(\pi)]\neg\mathbf{Com}(\pi)$, and the fact that commitment is a precondition for hope. Note however, that it is possible for an agent to become committed to a *new instance* of π and experience 'renewed hope.'

$\models \mathbf{fear}(\pi, \neg\kappa) \rightarrow [\mathbf{do}(\pi)]\neg\mathbf{fear}(\pi, \neg\kappa)$ (9)

Similarly to hope, fear only lasts for the duration of the prospect. Indeed, this follows directly from the corresponding property of **hope** above and the fact that hope is a precondition for fear. Note that this proposition does not say anything about whether or not the agent succeeded in bringing about κ by performing π, only that it will not stay afraid afterwards.

$\models \mathbf{B}[\mathbf{do}(\pi)]\neg\kappa \rightarrow (\neg\mathbf{hope}(\pi, \kappa) \wedge \neg\mathbf{fear}(\pi, \neg\kappa))$ (10)

If the agent believes it has no chance of accomplishing goal κ using plan π, then it will not hope for the impossible, nor fear the inevitable. The fact that there is also no fear in the consequent follows from the definition of **fear**, which validates $\models \neg\mathbf{hope}(\pi, \kappa) \rightarrow \neg\mathbf{fear}(\pi, \neg\kappa)$.

$\models \mathbf{fear}(\pi, \neg\kappa) \rightarrow \mathbf{B}(\langle\mathbf{do}(\pi)\rangle\kappa \wedge \langle\mathbf{do}(\pi)\rangle\neg\kappa)$ (11)

An agent experiencing fear with respect to a plan π and a goal κ believes that both the success and failure to accomplish κ are possible outcomes of performing π. This logical model does not tell anything about the likelihood of any of these outcomes; this is indeed something that will have to be taken into account by the quantitative model.

$\models \langle\mathbf{do}(\pi)\rangle\varphi \rightarrow [\mathbf{do}(\pi)]\varphi \quad \Rightarrow \quad \models \neg\mathbf{fear}(\pi, \neg\kappa)$ (12)

An agent that can predict the exact outcome of its actions will never experience fear. So with our definitions of **hope** and **fear**, we can express both the complementary nature of these emotions as described by OCC and the non-occurrence of fear in deterministic environments; in other words, an agent will never experience fear with respect to deterministic plans! This agrees with the intuitive notion that agents should not fear an undesirable outcome of deterministic actions, because they can predict the exact outcome beforehand. On the other hand, agents in nondeterministic or partially observable environments should always hope and fear simultaneously, because they cannot predict with absolute certainty whether or not their plan π for achieving goal κ will succeed.

6. EFFECTS ON DELIBERATION

Now that we have specified *when* an agent experiences hope or fear, we can try to specify *what to do* with these emotions. Recall that hope and fear are complementary emotions, so the specifications of the effects of these emotions must combine them both. There are two possible combinations of **hope** and **fear**:

1. **hope** but no **fear**: this case is similar to the effect of being *happy* as defined by Meyer & Dastani [8, 4]. When an agent is hopeful with respect to a plan and a goal, it is said by OCC to *be pleased* about how its plan is progressing, so it should keep its intention and commitment with respect to the plan and the goal. So in this case, the heuristic is to ensure that further deliberation of the agent, denoted as *deliberate*, does not change this:

$$(\mathbf{hope}(\pi, \kappa) \wedge \neg\mathbf{fear}(\pi, \neg\kappa)) \rightarrow [\mathbf{do}(\mathit{deliberate})](I(\pi, \kappa) \wedge \mathbf{Com}(\pi))$$

2. Simultaneous **hope** and **fear**: this is the more interesting case. The OCC model defines fear as *being displeased* about the prospect of an undesirable event. But what could the effect of being displeased be? Doing something about it! An agent experiencing fear with respect to a plan and a goal should be allowed to replan in order to find a new plan that can accomplish the goal:

$$(\mathbf{hope}(\pi, \kappa) \wedge \mathbf{fear}(\pi, \neg\kappa)) \rightarrow [\mathbf{do}(\mathit{deliberate})](\mathbf{Com}(\pi) \vee \mathbf{Com}(\mathit{replan}(\pi, \kappa, \pi'); \pi'))$$

We assume an agent has the ability to replan by performing the deliberation operation $\mathit{replan}(\pi, \kappa, \pi')$, which provides an alternative plan π' to achieve goal κ. Plan π' may depend on the original plan π or even be equal to π if an alternative plan cannot be found. For a proper and complete definition of the *replan* function, we refer the reader to Dastani et al. [3].

Is the formula above a good heuristic? No, because it is not specific enough. The $\mathbf{Com}(\pi) \vee \mathbf{Com}(\mathit{replan}(\pi, \kappa, \pi'); \pi')$ term does not specify *when* an agent should start replanning, only that it *may* do so. Because hope and fear are complementary emotions (i.e. their intensities always add up to a constant), a reasonable heuristic would be one that states that the agent should start replanning as soon as the intensity of the fear w.r.t. the undesirable event is greater than the intensity of the hope w.r.t. the desirable event. However, this cannot be expressed in a purely logical model. Therefore, a quantitative model of emotions is needed in order to complete the heuristic.

According to the OCC model, the variables affecting the intensity of hope w.r.t. a desirable event are 1) the degree to which the event is desirable and 2) the likelihood of the event. Analogously, the variables affecting the intensity of fear w.r.t. an undesirable event are 1) the degree to which the event is undesirable and 2) the likelihood of the event. We can thus define the *intensity* function for **hope** and **fear** as follows:

$$\begin{aligned} &\mathit{intensity}(\mathbf{hope}(\pi,\kappa)) := \\ &\qquad \mathcal{I}_{hope}(\mathit{desirability}(\kappa), \mathit{likelihood}(\pi,\kappa)), \\ &\mathit{intensity}(\mathbf{fear}(\pi,\neg\kappa)) := \\ &\qquad \mathcal{I}_{fear}(\mathit{undesirability}(\neg\kappa), \mathit{likelihood}(\pi,\neg\kappa)). \end{aligned}$$

Here the functions *desirability*, *undesirability*, and *likelihood* return application-dependent and platform-dependent measures of the (un)desirability of (not) achieving κ and the likelihood of the execution of plan π resulting in a state where κ holds, respectively. The functions $\mathcal{I}_{hope}$ and $\mathcal{I}_{fear}$ then combine these measures, according to this greatly simplified quantitative model, to a value in the range of $\{low, high\}$. It should be noted that the functions $\mathcal{I}_{hope}$, $\mathcal{I}_{fear}$, *desirability*, *undesirability*, and *likelihood* all implicitly depend on the agent and model–state pair passed to the *intensity* function (see Definition 3.8). Because hope and fear are complementary emotions, we assume that the $\mathcal{I}_{hope}$ and $\mathcal{I}_{fear}$ functions are constrained such that the intensity of hope is high if and only if the intensity of fear is low, i.e. that the following formulas define constraints on the agent model:

$$(\mathbf{hope}(\pi,\kappa) \wedge \neg\mathbf{fear}(\pi,\neg\kappa)) \rightarrow \mathit{intensity}(\mathbf{hope}(\pi,\kappa)) = high \quad (13a)$$

$$(\mathbf{hope}(\pi,\kappa) \wedge \mathbf{fear}(\pi,\neg\kappa)) \rightarrow (\mathit{intensity}(\mathbf{hope}(\pi,\kappa)) = high \leftrightarrow \mathit{intensity}(\mathbf{fear}(\pi,\neg\kappa)) = low) \quad (13b)$$

Now we can complete the heuristic by specifying that an agent should keep its commitment with respect to a plan while its hope with respect to that plan is greater than its fear, whereas the agent should start replanning when its fear is greater than its hope:

$$\begin{aligned} &(\mathbf{hope}(\pi,\kappa) \wedge \mathbf{fear}(\pi,\neg\kappa)) \rightarrow \\ &\quad ((\mathit{intensity}(\mathbf{hope}(\pi,\kappa)) = high \rightarrow \\ &\qquad [\mathbf{do}(\mathit{deliberate})](I(\pi,\kappa) \wedge \mathbf{Com}(\pi))) \wedge \\ &\quad (\mathit{intensity}(\mathbf{fear}(\pi,\neg\kappa)) = high \rightarrow \\ &\qquad [\mathbf{do}(\mathit{deliberate})](I(\pi'',\kappa) \wedge \mathbf{Com}(\pi'')))) \end{aligned}$$

where $\pi'' = (\mathit{replan}(\pi,\kappa,\pi');\pi')$.

7. CONCLUSION AND FUTURE WORK

We have identified three problems in multi-agent systems, namely nondeterminism in autonomous decision making, cooperation and coordination in multi-agent systems, and believability in human-agent interaction. Taking a functional approach, we have presented part of a formalization of the OCC model of secondary emotions. We expect that the presented affective model based on human-inspired emotions, once completed, provides a reasonable and useful solution to the mentioned problems. We have shown how the emotions hope and fear as described in the OCC model can influence the deliberation process of an agent, resulting in a decrease in nondeterminism in its decision making process.

At the time of writing, we have completed a logical (qualitative) formalization of all 22 emotions in the OCC model. We are currently working on a quantitative model incorporating emotions potentials, thresholds, and intensities, as well as investigating how functions like *desirability* and *likelihood* can be defined. For future work, we intend to develop transition semantics and an implementation of our formalization in an agent-oriented programming language. We will evaluate the emotional model by running it on the iCat (see Figure 1), which is a cat-shaped robot with a human-like face capable of making believable emotional expressions. Current scenarios for the robot are that of a companion robot for elderly and a cooking assistant.

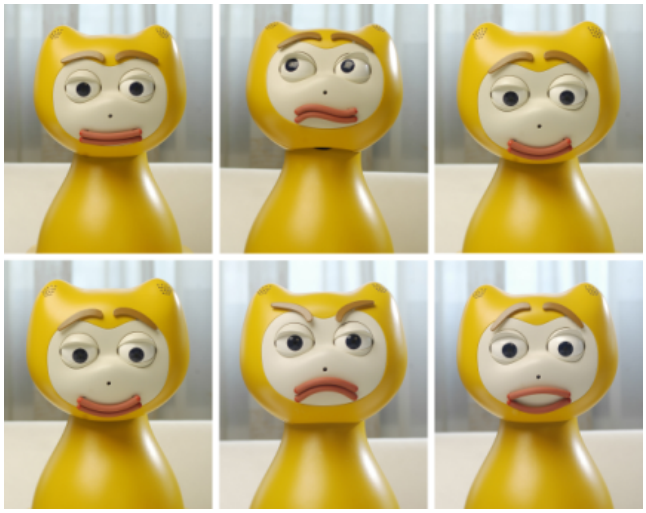

Figure 1: The iCat robot, developed by Philips, is an experimentation platform for human-robot interaction. See `http://www.research.philips.com/robotics`.

8. REFERENCES

[1] J. Bates, A. B. Loyall, and W. S. Reilly. An architecture for action, emotion, and social behaviour. In C. Castelfranchi and E. Wemer, editors, *Artificial Social Systems, MAAMAW-92 (LNAI Volume 830)*, pages 55–68, 1994.

[2] A. R. Damasio. *Descartes' Error: Emotion, Reason and the Human Brain*. Grosset/Putnam, New York, 1994.

[3] M. Dastani, F. S. de Boer, F. Dignum, and J.-J. C. Meyer. Programming agent deliberation: An approach illustrated using the 3APL language. In J. S. Rosenschein, T. Sandholm, M. Wooldridge, and M. Yokoo, editors, *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'03)*, pages 97–104, Melbourne, Australia, 2003. ACM Press.

[4] M. Dastani and J.-J. C. Meyer. Programming agents with emotions. In *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI'06)*, 2006.

[5] C. Elliott. Components of two-way emotion communciation between humans and computers using a broad, rudimentary, model of affect and personality. In *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, volume 1(2):16–30. 1994.

[6] C. Elliott. I picked up catapia and other stories: A multimodal approach to expressivity for "emotionally intelligent" agents. In *Proceedings of the First International Conference on Autonomous Agents*, pages 451–457, 1997.
[7] T. Koda. Agents with faces: A study on the effects of personification of software agents. Master's thesis, MIT Media Laboratory, 1996.
[8] J.-J. C. Meyer. Reasoning about emotional agents. In R. L. de Mántaras and L. Saitta, editors, *Proceedings of the 16th Eureopean Conference on Artificial Intelligence (ECAI'04)*, pages 129–133. IOS Press, 2004.
[9] J.-J. C. Meyer, W. van der Hoek, and B. van Linder. A logical approach to the dynamics of commitments. *Arificial Intelligence*, 113:1–40, 1999.
[10] W. S. Neal Reilly. *Believable Social and Emotional Agents*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, 1996.
[11] K. Oatley and J. M. Jenkins. *Understanding Emotions*. Blackwell Publishing, Oxford, UK, 1996.
[12] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge, UK, 1988.
[13] R. W. Picard. *Affective Computing*. MIT Press, 1997.
[14] A. Sloman. Beyond shalow models of emotion. *Cognitive Processing*, 2(1):177–198, 2001.

APPENDIX

A. PROOFS OF PROPOSITIONS

The belief and action modalities are closed under implication, so we use the common $\Box$-axioms for $\Box \in \{\mathbf{B}, [\mathbf{do}(\alpha)]\}$, i.e. $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$. In our proofs there is often a condition χ for $\Box\varphi$, in which case we use the axiom in the following form: $(\chi \to \Box\varphi) \wedge \Box(\varphi \to \psi) \to (\chi \to \Box\psi)$.

PROOF. (Proposition (8)).

1. $\mathbf{hope}(\pi,\kappa) \to \mathbf{Com}(\pi)$	From Def. (6)
2. $\mathbf{Com}(\pi) \to [\mathbf{do}(\pi)]\neg\mathbf{Com}(\pi)$	Constraint (4)
3. $\mathbf{hope}(\pi,\kappa) \to [\mathbf{do}(\pi)]\neg\mathbf{Com}(\pi)$	1, 2, HS[1]
4. $\neg\mathbf{Com}(\pi) \to \neg\mathbf{hope}(\pi,\kappa)$	1, Contraposition
5. $[\mathbf{do}(\pi)](\neg\mathbf{Com}(\pi) \to \neg\mathbf{hope}(\pi,\kappa))$	4, Necessitation
6. $\mathbf{hope}(\pi,\kappa) \to [\mathbf{do}(\pi)]\neg\mathbf{hope}(\pi,\kappa)$	3, 5, $\Box$-axiom

□

PROOF. (Proposition (9)).

1. $\mathbf{fear}(\pi,\neg\kappa) \to \mathbf{hope}(\pi,\kappa)$	From Def. (7)
2. $\mathbf{hope}(\pi,\kappa) \to [\mathbf{do}(\pi)]\neg\mathbf{hope}(\pi,\kappa)$	Prop. (8)
3. $\mathbf{fear}(\pi,\neg\kappa) \to [\mathbf{do}(\pi)]\neg\mathbf{hope}(\pi,\kappa)$	1, 2, HS
4. $\neg\mathbf{hope}(\pi,\kappa) \to \neg\mathbf{fear}(\pi,\neg\kappa)$	1, Contrapos.
5. $[\mathbf{do}(\pi)](\neg\mathbf{hope}(\pi,\kappa) \to \neg\mathbf{fear}(\pi,\neg\kappa))$	4, Necessitation
6. $\mathbf{fear}(\pi,\neg\kappa) \to [\mathbf{do}(\pi)]\neg\mathbf{fear}(\pi,\neg\kappa)$	3, 5, $\Box$-axiom

□

[1]HS stands for hypothetical syllogism.

PROOF. (Proposition (10)).

1. $\mathbf{B}[\mathbf{do}(\pi)]\neg\kappa \leftrightarrow \mathbf{B}\neg\langle\mathbf{do}(\pi)\rangle\kappa$	$\langle\cdot\rangle = \neg[\cdot]\neg$
2. $\mathbf{B}\neg\langle\mathbf{do}(\pi)\rangle\kappa \to \neg\mathbf{B}\langle\mathbf{do}(\pi)\rangle\kappa$	Seriality of $\mathbf{B}$
3. $\neg\mathbf{B}\langle\mathbf{do}(\pi)\rangle\kappa \to \neg I(\pi,\kappa)$	From Definition (5)
4. $\neg I(\pi,\kappa) \to \neg\mathbf{hope}(\pi,\kappa)$	From Definition (6)
5. $\neg\mathbf{hope}(\pi,\kappa) \to \neg\mathbf{fear}(\pi,\neg\kappa)$	From Definition (7)
6. $\mathbf{B}[\mathbf{do}(\pi)]\neg\kappa \to \neg\mathbf{hope}(\pi,\kappa)$	1–4, HS
7. $\mathbf{B}[\mathbf{do}(\pi)]\neg\kappa \to \neg\mathbf{fear}(\pi,\neg\kappa)$	1–5, HS
8. $\mathbf{B}[\mathbf{do}(\pi)]\neg\kappa \to (\neg\mathbf{hope}(\pi,\kappa) \wedge \neg\mathbf{fear}(\pi,\neg\kappa))$	6, 7, Combined

□

PROOF. (Proposition (11)).

1. $\mathbf{fear}(\pi,\neg\kappa) \to \mathbf{hope}(\pi,\kappa)$	From Definition (7)
2. $\mathbf{hope}(\pi,\kappa) \to \mathbf{B}\langle\mathbf{do}(\pi)\rangle\kappa$	From Definition (6)
3. $\mathbf{fear}(\pi,\neg\kappa) \to \mathbf{B}\langle\mathbf{do}(\pi)\rangle\kappa$	1, 2, HS
4. $\mathbf{fear}(\pi,\neg\kappa) \to \mathbf{B}\langle\mathbf{do}(\pi)\rangle\neg\kappa$	From Definition (7)
5. $\mathbf{fear}(\pi,\neg\kappa) \to (\mathbf{B}\langle\mathbf{do}(\pi)\rangle\kappa \wedge \mathbf{B}\langle\mathbf{do}(\pi)\rangle\neg\kappa)$	3, 4, Combined

□

PROOF. (Proposition (12)). If we consider deterministic actions (or plans) π, we have the property that all possible executions of π produce the same result, i.e. $\models \langle\mathbf{do}(\pi)\rangle\varphi \to [\mathbf{do}(\pi)]\varphi$. We have seen in Proposition (11) that $\mathbf{fear}(\pi,\neg\kappa)$ implies $\mathbf{B}(\langle\mathbf{do}(\pi)\rangle\kappa \wedge \langle\mathbf{do}(\pi)\rangle\neg\kappa)$. However, for deterministic π, this implies $\mathbf{B}([\mathbf{do}(\pi)]\kappa \wedge [\mathbf{do}(\pi)]\neg\kappa)$, which implies $\mathbf{B}[\mathbf{do}(\pi)]\bot$. This in turn contradicts the condition for **fear** that $\mathbf{B}\langle\mathbf{do}(\pi)\rangle\neg\kappa$ must hold. Consequently, $\mathbf{fear}(\pi,\neg\kappa)$ implies $\neg\mathbf{fear}(\pi,\neg\kappa)$; in other words, $\models \neg\mathbf{fear}(\pi,\neg\kappa)$. □