Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 15–16 (2012) 33–49

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

# Coppélius' concoction: Similarity and complementarity among three affect-related agent models

Action Editor: Ron Sun

Johan F. Hoorn, Matthijs Pontier *, Ghazanfar F. Siddiqui

*Center for Advanced Media Research Amsterdam (CAMeRA@VU), VU University Amsterdam, De Boelelaan 1081, 1081HV Amsterdam, The Netherlands*

Received 14 September 2010; received in revised form 27 April 2011; accepted 30 April 2011
Available online 27 May 2011

**Abstract**

In aiming for behavioral fidelity, artificial intelligence cannot and no longer ignores the formalization of human affect. Affect modeling plays a vital role in faithfully simulating human emotion and in emotionally-evocative technology that aims at being real. This paper offers a short expose about three models concerning the regulation and generation of affect: CoMERG, EMA and I-PEFiC$^{ADM}$, which each in their own right are successfully applied in the agent and robot domain. We argue that the three models partly overlap and where distinct, they complement one another. To enable their integration, we provide an analysis of the theoretical concepts, resulting in a more precise representation of affect simulation in virtual humans, which we verify with simulation tests.


## 1. Introduction

Recently, much research has been dedicated to developing more realistic Intelligent Virtual Agents (IVAs). However, these agents are often emotionally not very human-like. For example, many IVAs can show emotions using facial expressions or the tone of their voice, but most of them still struggle to be believable in terms of what emotions to show at what moment (e.g., emotion regulation (Marsella & Gratch, 2003), stress and work-load (Endsley, 1995), and moods (Beck, 1987)); let alone that they actually understand and react empathically to the emotional state of other agents or human users. Previous research has shown that closely mimicking humans is important for an agent to increase human involvement in a virtual environment (e.g., Van Vugt, Hoorn, Konijn, & De Bie Dimitriadou, 2006), although mimicking is not always effective; (e.g., Ekman & Rosenberg, 2005).

To create a more natural communication system in intelligent agents, our earlier work and that of our colleagues in the Institute for Creative Technologies, University of Southern California focused on different aspects of emotion generation, regulation, and affective processes. Marsella and Gratch (2006, 2009) formalized the theory of emotion and adaptation of Smith and Lazarus (1990) into *EMA* to create agents that cope with negative affect. The emotion-regulation theory of Gross (2001) inspired Bosse, Pontier, and Treur (2007) to develop CoMERG (the Cognitive Model for Emotion Regulation based on Gross). Hoorn, Pontier, and Siddiqui (2008) used the concern-driven theory of Frijda (1986) to design I-PEFiC$^{ADM}$ and built agents that can trade rational for affective choices.

Together, these theories cover a large part of appraisal-based emotion theory (Frijda, Smith & Lazarus, Gross) and all three boil down to appraisal models of emotion. We therefore expected that the related computational models would nicely fit together so that we would better account for the complexity of human behavioral affect than the separate approaches would do alone.

All three approaches point at important aspects of human affective behavior, but each misses out on some aspect. CoMERG and EMA address the regulation of affective states, but Gratch and Marsella (2006) Marsella and Gratch (2009) do not regulate positive affect.

* Corresponding author.
*E-mail addresses:* jf.hoorn@camera.vunl (J.F. Hoorn), mpr210@few.vu.nl (M. Pontier), ghazanfa@few.vu.nl (G.F. Siddiqui).

1389-0417/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2011.04.001

CoMERG has no provisions for generating affect and does not account for a causal interpretation of the world-state. I-PEFiC$^{ADM}$ (Hoorn et al., 2008) generates and balances affect but addresses only one regulation mechanism. We created a new model, Silicon Coppélia, in which we combined the most useful parts of the three models mentioned above. This integrated model is expected to simulate richer agent behavior than what CoMERG, EMA and I-PEFiC$^{ADM}$ can do alone. We will test this by performing simulation experiments on Silicon Coppélia under various parameter settings.

We coined our integrated and simulated model “Silicon Coppélia” after the mechanic doll, dancing in the ballet of Arthur Saint-Léon (1870). She is the concoction of sinister Doctor Coppélius, who made her so human-like that a young man named Franz was prepared to denounce his fiancée Swanilda for her.

## 2. Model overview

We collected three models (CoMERG, EMA, and I-PEFiC$^{ADM}$) of agent affect-generation and affect-regulation. In our view these models offer plenty of opportunities for integration.

Note that the presented models embody a particular variant of an affect theory in that they have some unique properties that distinguish them from their original sources. Many design choices underlying such models arise from the need to create a working computational system, a challenge the original theorists have never confronted.

### 2.1. CoMERG

Gross (2001) states that “Emotion regulation includes all of the conscious and unconscious strategies we use to increase, maintain, or decrease one or more components of an emotional response.” Bosse, Pontier, Siddiqui, and Treur (2007) developed difference equations and logical rules to simulate the dynamics of Gross’ emotion-regulation strategies. The CoMERG model was incorporated into agents in a virtual storytelling application (Bosse et al., 2007). Gross distinguishes five different emotion-regulation strategies: *situation selection, situation modification, attentional deployment, cognitive change* and *response modulation* (see Fig. 1). Humans have strategies to influence the level of emotions to avoid extreme responses.

Gross distinguishes (1) an *experiential* component (subjective feeling), (2) a *behavioral* component (behavioral responses), and (3) a *physiological* component (e.g., heartbeat and respiration). In *situation selection*, a person chooses the situation that matches the preferred emotional response level (e.g., a person may not want to go to a party because (s)he dislikes someone who will go there too). In *situation modification*, a person changes the situation to obtain a different level of emotion (e.g., zapping to another channel because of an annoying performer).

Attentional deployment refers to shifting focus (e.g., when looking away from a scary movie scene). *Cognitive change* selects a cognitive meaning to an event (e.g., the weather is blamed for losing a match). With *response modulation*, people try to influence the process that response tendencies may become a behavioral response (e.g., hiding shyness).

### 2.2. Emotion and adaption (EMA) model

EMA (Fig. 2) is a computational model of the cognitive antecedents and consequences of emotion as posited by Smith and Lazarus (1990). In cognitive appraisal theories, appraisal and coping center on people’s *interpretation* of

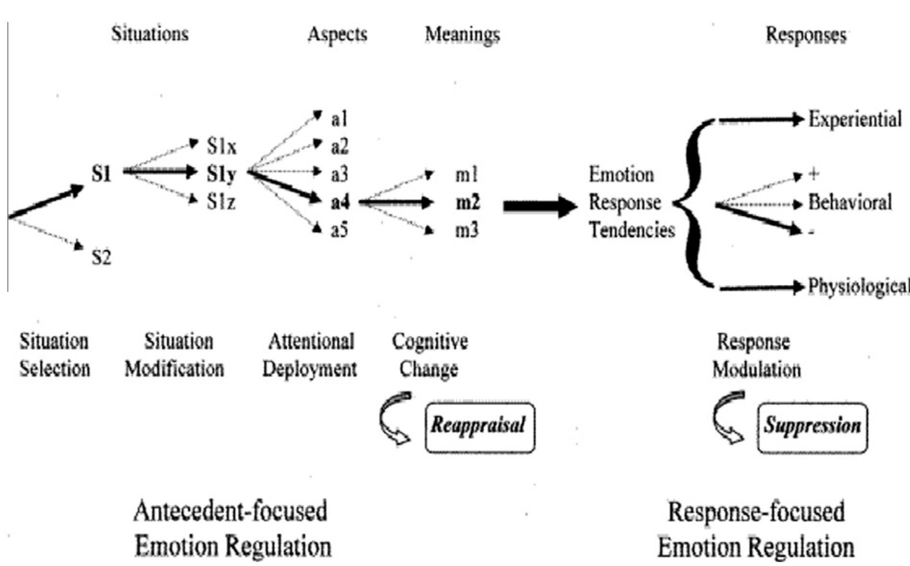



Fig. 1. Emotion regulation model by Gross (2001).

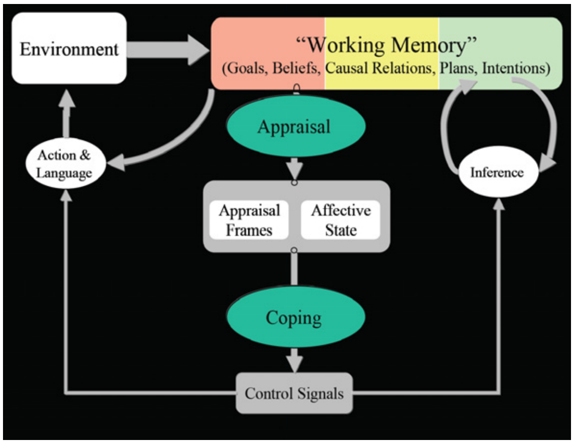



Fig. 2. Computational instantiation of the cognitive–motivational–emotive system.

their relationship with the environment. This interpretation is constructed by cognitive processes, summarized by appraisal variables, and altered by coping responses. EMA maintains an explicit symbolic representation of the relationship between events and an agent's internal beliefs, desires, and intentions. AI planning then makes explicit use of such event-consequence representations, while a BDI framework handles the epistemic factors (i.e. beliefs and intentions) that underlie social activities.

Appraisal processes characterize this representation in terms of individual appraisal judgments. These extend traditional AI concerns such as the following, with notions of utility and probability, e.g., desirability, likelihood, causal attribution, controllability and changeability.

Patterns of appraisal elicit emotional displays but they also initiate coping processes to regulate the agent's cognitive response to the appraised emotion. Coping strategies work in the reverse direction of appraisal, identifying plans, beliefs, desires or intentions to maintain or alter in order to maintain positive or reduce negative emotional appraisals. These include "problem focused" strategies (e.g., planning) directed towards improving the world (the traditional concern of AI techniques) but also encompasses "emotion-focused" strategies that impact an agent's epistemic and motivational state. Some examples of copingstrategies are planning, seeking instrumental support, procastrination, denial, mental disengagement and blame shifting.

The above strategies provide input to the cognitive processes that actually execute these directives. For example, planful coping generates an intention to act, leading a planning system associated with EMA to generate and execute a valid plan to accomplish this act. Alternatively, coping strategies might abandon the goal, lower the goal's importance, or re-assess who is to blame.

EMA is a fully implemented model and has been applied to simulate realistic human emotional responses (e.g., decision-making and nonverbal behaviors of computer-generated role-players in a variety of social training environments). Several empirical studies demonstrated EMA's effectiveness in modeling the influence of emotion over human judgments when compared with human behavior in laboratory settings (Mao & Gratch, 2006).

The algorithm of EMA is as follows (Marsella & Gratch, 2009):

1. Construct and maintain a causal interpretation of ongoing world events in terms of beliefs, desires, plans, and intentions.
2. Generate multiple appraisal frames that characterize features of the causal interpretation in terms of appraisal variables.
3. Map individual appraisal frames into individual instances of emotion.
4. Aggregate emotion instances into a current emotional state and overall mood.
5. Adopt a coping strategy in response to the current emotional state.

### 2.3. I-PEFiC$^{ADM}$

Originally, the empirically validated framework of Perceiving and Experiencing Fictional Characters (PEFiC) described the receiver's reception of fictional characters (Konijn & Hoorn, 2005). Later versions were applied to embodied agents with user interaction possibilities, resulting into Interactive PEFiC (Van Vugt, Hoorn, & Konijn, 2009). I-PEFiC was then used to model affective behavior of agents, while a module for Affective Decision Making was added to simulate irrational agent behavior, hence I-PEFiC$^{ADM}$ (Hoorn et al., 2008). I-PEFiC$^{ADM}$ assumes an encoding, a comparison, and a response phase (Fig. 3). During *encoding*, the agent perceives the user and the situation the user is in, in terms of *ethics*, *aesthetics*, *epistemics* and *affordances*.. Affordances are action possibilities that make the user instrumental to achieve agent goals (e.g., maintenance, security).

In the *comparison* phase, the agent appraises the *relevance* and the valence of user features to agent goals. Relevance determines the intensity of the effect $[0, 1]$, while valence determines its direction $[-1, 1]$. User features may afford the facilitation of a desired agent goal. Additionally, the agent estimates a level of *similarity* between agent and user.

In the *response* phase, the agent establishes the levels of *involvement* with and *distance* towards the user. These two tendencies occur in parallel and compensate one another. In addition, the agent calculates a value for the so called *use intentions*, the willingness to employ the user again to achieve agent goals.

The affective decision making manages that the agent makes a decision on the more rationally generated use intentions in unison with the more affectively generated involvement-distance trade-off. This enables the agent to make irrational choices where this is considered to be human-like.

## 3. Triple comparison

Fig. 4 depicts the similarities and differences between CoMERG, EMA, and I-PEFiC$^{ADM}$. By and large, all three assume the perception of situational features that provoke subsequent appraisals, which are related or matched to goals and desires. All three describe affective responses (overt or covert) and the regulation of those responses. Next, we offer a comparison of models, using Fig. 4 as our central reference point.

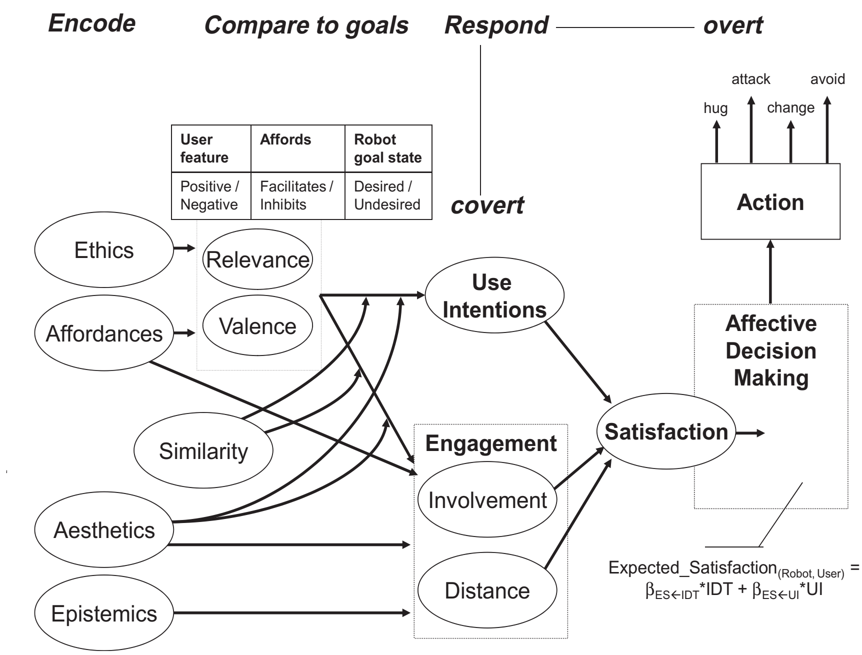
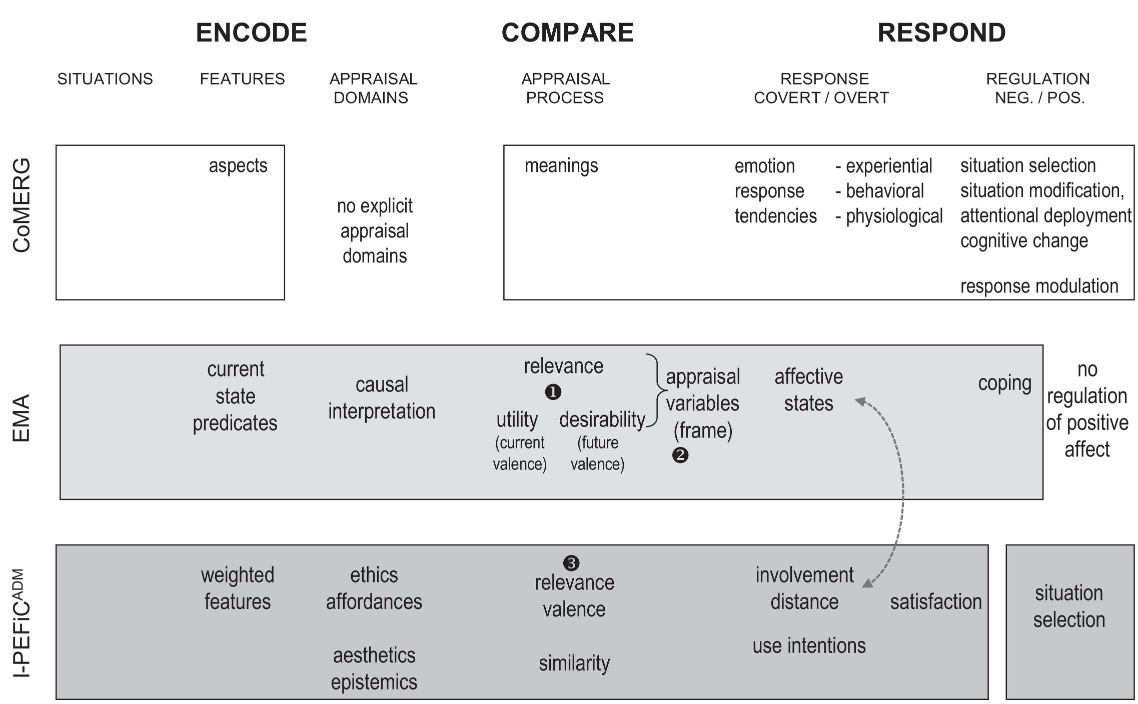



Fig. 3. Dependencies in I-PEFiCADM. Curved arrows indicate interaction effects. IDT = involvement-distance trade-off, UI = use intentions.




❶ Current state predicates ('features') are matched against goals etc. of self and others (allows taking perspectives)
❷ Appraisal frames are specifications of judgments on relevance and 'valence' (i.e. utility or desirability)
❸ Features matched against goals etc. of self (does not account for change in perspective)

Fig. 4. Graphical overview of CoMERG, EMA, and I-PEFiC$^{ADM}$.

Features in CoMERG are called “aspects” and people can focus on different aspects (features) of the world to regulate emotions. EMA does not work with ‘features’ as such but the “current state predicates” are statements about features of the environment that can be true or false. In I-PEFiC$^{ADM}$, features receive a certain weight according to frequency of occurrence, salience, or familiarity. What exactly the weight should indicate is empirically an unsettled matter. Weights can change as a function of attentional shifts, switching foci, or situation changes.

With respect to *appraisal domains*, CoMERG does not explicitly mention any, whereas PEFiC (e.g., Konijn & Hoorn, 2005) and I-PEFiC (e.g., Van Vugt et al., 2009) focus their appraisal domain on perceiving characters. For the judgment of fictional characters and embodied agents, users classify features as good or bad, beautiful or ugly, realistic or unrealistic, and as aids or obstacles (Van Vugt et al., 2009). For user engagement and use intentions, appraisals of ethics and affordances are the most important determinants. Appraisal domains in EMA focus on interpreting the relationship with the environment. In EMA, agents perceive the world according to a causal interpretation of past and ongoing events, including past actions and plans as well as intentions of self and others.

### 3.1. Compare

The *appraisal process* is least explained in CoMERG, which uses “meanings”. As an emotion-regulation strategy in CoMERG, people can attach different cognitive meaning to a situation. One type of such ‘cognitive change’ is called *reappraisal* (Gross, 2001), which is a re-evaluation of a potentially emotion-eliciting situation to decrease emotional impact (e.g., poor personal performance is blamed on the weather). I-PEFiC$^{ADM}$ attaches personal meaning to a feature through relevance and valence. Something may potentially benefit or harm goals, beliefs, or concerns and as such, acquires individual ‘meaning’ (Frijda, 1986; Frijda, 1988, cf. primary appraisal in Lazarus, 1991).

In EMA, this meaning is acquired through multiple appraisal frames, which allow for perspective taking. Appraisal frames are generated from many appraisal variables, which are called appraisal components by Smith and Lazarus and appraisal dimensions by Roseman (1984). Most of these appraisal variables can be mapped to relevance and valence as used in I-PEFiC$^{ADM}$. According to EMA, ‘relevance’ measures the significance of an event for the agent. Unlike Frijda, however, EMA equates significance with utility, which in Frijda’s terms would be ‘valence.’ An event outcome is only deemed significant in EMA if it facilitates or inhibits a state predicate with non-zero utility. Valence is not explicitly mentioned in EMA although “utility” and “desirability” can be regarded as two instantiations of it.

Utility is a measure of the relative satisfaction from (or desirability of) environmental features. EMA represents preferences over environmental features as numeric utility over the truth-value of state predicates. Utilities may be either intrinsic (meaning that the agent assigns intrinsic worth to this environmental feature) or extrinsic (meaning that they inherit worth through their probabilistic contribution to an intrinsically valuable state feature). Utility, then, may be viewed as positive or negative outcome expectations about features in the current situation and is expressed in current state predicates (hence, ‘current valence’).

Desirability covers both a notion of intrinsic pleasantness and goal congruence (in Scherer’s (1993) typography), as well as a measure of importance or relevance. It captures the appraised valence of an event with regard to an agent’s preferences. An event is desirable, from some agent’s perspective, if it facilitates a state to which the agent attributes positive utility or if it inhibits a state with negative utility. An event is undesirable if it facilitates a state with negative utility or if it inhibits a state with positive utility. Like utility, desirability may be viewed as positive or negative outcome expectations but this time about features in the future situation (‘future valence’).

The explicit division in current and future states is what I-PEFiC$^{ADM}$ is missing as well as the possibility to change perspectives. EMA and I-PEFiC$^{ADM}$ resemble each other in that causal interpretation of ongoing world events in terms of beliefs, desires, plans, and intentions in EMA is comprised in the beliefs, goals, and concerns that are checked for relevance and valence in I-PEFiC$^{ADM}$. However, EMA uses a number of variables, called appraisal frames, to cover the appraisal process, whereas in I-PEFiC$^{ADM}$, these appraisal frames appear to pertain to the more general concepts of relevance and valence. For example, urgency would be a clear-cut specification of relevance (cf. Frijda, 1988, p. 352) and ego involvement could be seen as a part of valence. However, EMA also uses some variables (such as causal attribution and coping potential) which are more related to the environment and less to the character, and which are somewhat broader than relevance and valence.

### 3.2. Respond

Fig. 4 exemplifies that in EMA, relevance of an event as well as utility and desirability (current and future valence) of features are mapped via an appraisal frame onto emotion instances of a particular category and intensity. These are called affective states. This may be seen as a covert response to the situation – an internal affective state that does not yet translate into overt actions. In I-PEFiC$^{ADM}$, affective states as such are not the focus, but rather the involvement-distance trade-off is, which is seen as the central process of engagement.

What comes closest to EMA’s affective states are involvement-distance in combination with what I-PEFiC$^{ADM}$ calls “emotions” (Fig. 3, curved arrows). On this view, emotions are byproducts of the trade-off. For example, when Franz woos Coppélia, her involvement with him may be accompanied by happiness. When he looks at another girl, Coppélia may still be involved with Franz but this time she may feel challenged.

The involvement-distance trade-off could also count as the concretization of the emotion-response tendencies that CoMERG hinges on. In CoMERG, these tendencies result into several responses: experiential, behavioral, and physiological. EMA and I-PEFiC$^{ADM}$ are restricted to the experiential and behavioral domain. In EMA, affective states (experiential) lead to coping behavior. Here, coping is instantiated within some specific domain action. For example, if Franz has pathological fear towards dating androids and suspects Coppélia of being one, he might adopt emotion-focused coping (e.g., engage in wishful thinking and lower the probability she is an android) which will inform the next decision; or he might adopt problem-focused coping to take a specific overt action to address the threat (i.e. destroy Coppélia). In I-PEFiC$^{ADM}$, the combination of involvement, distance, and use intentions predicates the level of satisfaction (experiential), which feeds into affective decision making. This results into overt responses (behavior) such as kissing, kicking, or walking away.

CoMERG proposes a broad model of five *emotion-regulation strategies* (see Section 2.1). Following Gross, CoMERG predicts that strategies that are performed earlier in the process are more effective to regulate one's emotions. EMA provides a more specific model which focuses (in much detail) on coping. Situation selection and situation modification are implemented in EMA via problem-focused coping strategies (i.e. take-action) and avoidance. The domain-model given to EMA must encode that the situational features producing a negative emotion would be reversed/blocked by the effects of some action. *Attentional deployment* corresponds to EMA's strategies to seek/suppress information. Cognitive change corresponds to EMA's various emotion-directed strategies. EMA does not model suppression. I-PEFiC$^{ADM}$ focuses on situation selection. Another difference is that CoMERG and I-PEFiC$^{ADM}$ allow the regulation of affect by increasing, maintaining, or decreasing the positive or negative response, whereas EMA focuses on decreasing negative affect alone. In EMA, being overenthusiastic is left uncontrolled, whereas in CoMERG and I-PEFiC$^{ADM}$, positive affect can be down-regulated or compensated for e.g., in CoMERG positive affect can be suppressed and negative affect can be up regulated. As a result, one can state that the coping in EMA is superceded by the emotion regulation in CoMERG.

For EMA, there must be an explicit causal connection between coping strategies and the emotions they are regulating whereas for CoMERG that is not a prerequisite. In CoMERG, people perform strategies to change their level of emotion but how this works is described informally. EMA gives a more detailed and formal description of how emotion regulation works. For example, reappraisal as a general emotion-regulation strategy in CoMERG is in EMA described in terms of a change in causal interpretation.

In EMA, several emotions are combined to calculate an 'overall mood'. I-PEFiC$^{ADM}$ supports several affective processes to be calculated at the same time but there is not a way to calculate an overall mood. I-PEFiC$^{ADM}$ focuses on the trade-off between involvement and distance and not on coping per se. Also in CoMERG, several emotion-regulation strategies may be performed simultaneously, but it is not mentioned how these aggregate into an 'overall mood'.

## 4. Conceptual decisions

We adhered to the linguistic convention that an agent detects 'features' in a situation instead of 'aspects', because both EMA and I- PEFiCADM use that concept and because it is interchangeable with "aspects" in CoMERG. Moreover, the term 'features' better fits mathematical approaches that use feature sets to calculate certain values for a situation.

Only I-PEFiC$^{ADM}$ explicitly mentions the appraisal domains that are important in perceiving features. Therefore, the agent will use ethics, affordances, aesthetics, and epistemics as the main domains through which features are funneled into the appraisal process.

CoMERG, EMA, and I-PEFiC$^{ADM}$ all assume or elaborate an appraisal process. CoMERG is least explicit and the concept of 'meaning' can easily be attached to 'personal significance' and 'personal relevance' in both EMA and I-PEFiC$^{ADM}$. In EMA and I-PEFiC$^{ADM}$, relevance and valence play an active role, but EMA models the specific manifestations rather than the general concepts. In unison, we will use the term relevance to indicate importance or meaning to personal goals, concerns, beliefs, intentions, plans, etc. and valence to indicate (current) utility or (future) desirability of features in a situation. This may instantiate in the form of, for example, urgency as an aspect of relevance and likelihood or unexpectedness as an aspect of future valence.

On the response side, EMA focuses on moods and emotions whereas I-PEFiC$^{ADM}$ emphasizes the more general trends of involvement, distance, and use intentions. Yet, they are two sides of the coin that could be called 'affective states'. Emotions and moods may evolve from involvement-distance trade-offs and both the specific (e.g., happy emotions) and general experiential response (e.g., involvement) may be liable to regulation strategies.

CoMERG provides the most profound distinctions with respect to the type of responses (experiential, behavioral, and physiological) and the number of regulation strategies. However, in no way are these distinctions at odds with EMA or I-PEFiC$^{ADM}$. Coping is best worked out by EMA and situation selection by I-PEFiC$^{ADM}$. The latter encompasses a module for affective decision making that on the basis of expected satisfaction chooses from four distinct kinds of overt behaviors (i.e., fight, flight, change, embrace). Note that strategies such as CoMERG's response modulation are on the response side of the affect process but that they impinge upon encoding aspects: the situations and features, respectively (see next).

## 5. Silicon Coppélia

Fig. 5 shows how we combined CoMERG, EMA, and I-PEFiC$^{ADM}$ into Silicon Coppélia, a framework for computerized affect generation and regulation. When on the far left of Fig. 5, Coppélia looks away from Franz, attentional deployment makes her weigh the features that make Franz attractive or not (e.g., dancing capacities over loyalty).

Coppélia develops state predicates about Franz – and her situation with him – upon which she will eventually base her decision to adore him or leave. Features receive indices for different appraisal domains. Franz acquires personal meaning or significance for Coppélia because she compares his features with her personal goals, beliefs, and concerns. This establishes Franz' relevance and valence to Coppélia. While relevance determines the intensity of affect, valence governs its direction. Coppélia can also take perspectives and look at Franz through the eyes of Swanilda.

With the appraisal process completed, Coppélia is ready to affectively respond to Franz. Relevance, current and future valence form an appraisal frame that feeds into her (un)willingness to use Franz for her purposes (e.g., dancing, getting married) and that helps her trade friendship (involvement) for keeping her cool (distance). Inside, Coppélia now experiences several (perhaps ambiguous) emotions and moods. On a physiological level, she may be aroused (e.g., increased brain activity). All this is not visible for Franz yet; they are covert responses.

During affective decision making, Coppélia selects the option that promises the highest expected satisfaction and chooses from four overt actions: positive approach (e.g., compliment Franz on his dancing), negative approach (e.g., push him away), change (e.g., teach Franz on moral standards), or walk away from him. She might do each action one at a time. This may be accompanied by physiological reactions such as blushing and trembling. Response modulation may influence the affective decision making.

## 6. Implementation

In this section, we discuss the implementation of Silicon Coppélia, focusing on the variables mentioned in the previous sections, and the agents' belief system, which lead the agent to 'experience' joy, distress, hope, fear, anger, guilt, and surprise.

### 6.1. Encoding phase

According to I-PEFiC$^{ADM}$, an agent perceives another agent in terms of ethics (good/bad), aesthetics (beautiful/ugly), affordances (aid/obstacle) and realism (cf. Van Vugt et al., 2009).

Each agent has a value for 'designed beautiful/ugly'. This is a value the designer expects to raise in the user,

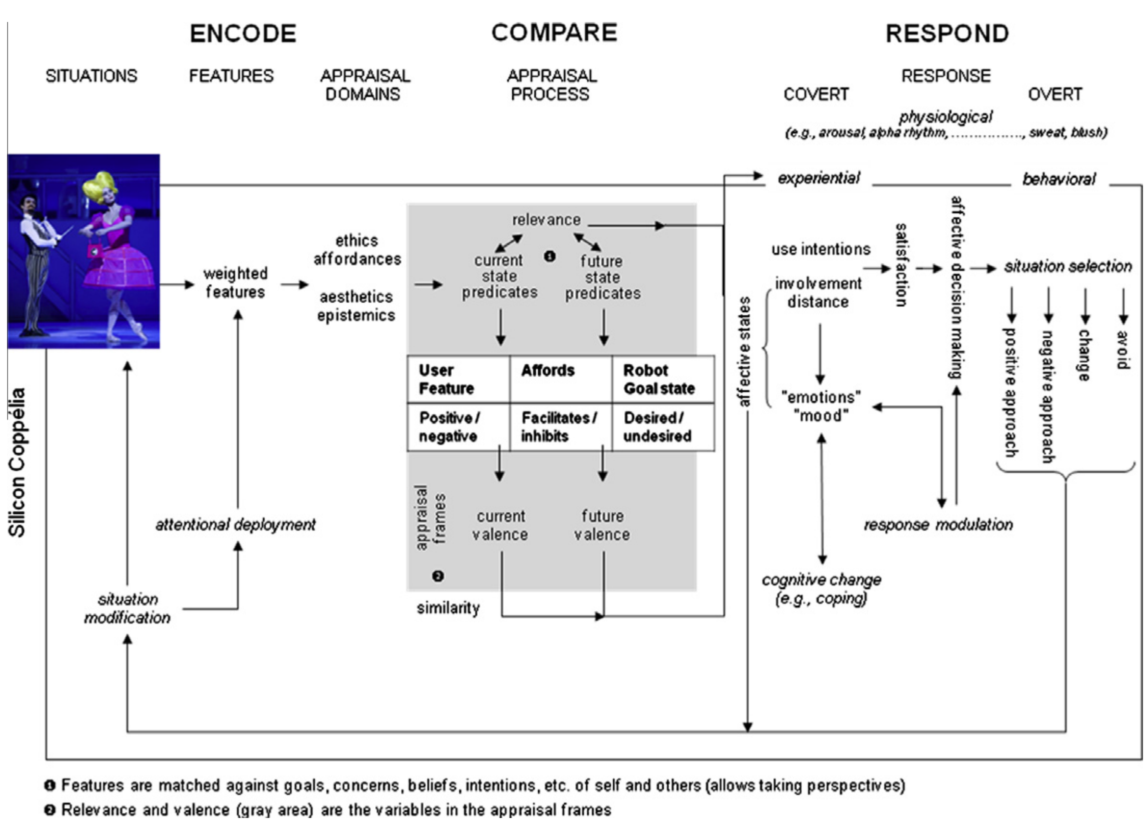



Fig. 5. Silicon Coppélia is the integration of CoMERG, EMA, and I-PEFiC$^{ADM}$.

or in another agent, based on general principles of aesthetics. This value could be seen as the mean 'score' an agent receives for its beauty/ugliness from all other agents. This designed value has a data-driven influence on how agents perceive the beauty of another agent. The variable bias represents the concept-driven influence on how agents perceive other agents' beauty. Depending on its value, a bias may increase or decrease an agent's perception of another agent's beauty. Note that 'another agent' could also be the agent itself. This is represented by the following formulas, given in mathematical format. In the formulas use in this document, $\text{Perceived}_{(<\text{Feat}>,A1,A2)}$ is agent A1's perception of agent A2's feature. Similarly, $\text{Bias}_{(A1,A2,\text{Beau})}$ is agent A1's bias in perceiving that feature.

$$\text{Perceived}_{(\text{Beau},A1,A2)} = \text{Bias}_{(A1,A2,\text{Beau})} * \text{Designed}_{(\text{Beau},\ A2)}$$
$$\text{Perceived}_{(\text{Ugly},A1,A2)} = \text{Bias}_{(A1,A2,\text{Ugly})} * \text{Designed}_{(\text{Ugly},\ A2)}$$

When two agents meet for the first time, they will assign a *perceived value* in the range [0, 1] to each other's features according to the formulas above. Bias in the range [0, 2] is multiplied with the designed value for the feature in the range [0, 1]. If agent A1 has a bias of 1 for, for instance, the beauty of agent A2, then A1 does not under- or overestimate the beauty of A2. If the bias is bigger than 1, then A1 is relatively positive about the beauty of agent A2. When the resulting value for the perceived feature is bigger than 1, it is set to 1, to prevent the formula from going out of range.

Similar formulas are used for ethics, epistemics and intended affordances:

$$\text{Perceived}_{(\text{Real},A1,A2)} = \text{Bias}_{(A1,A2,\text{Realistic})} * \text{Designed}_{(\text{Realistic},\ A2)}$$
$$\text{Perceived}_{(\text{Unr},A1,A2)} = \text{Bias}_{(A1,A2,\text{Unr})} * \text{Designed}_{(\text{Unr},\ A2)}$$
$$\text{Perceived}_{(\text{Good},A1,A2)} = \text{Bias}_{(A1,A2,\text{Good})} * \text{Designed}_{(\text{Good},\ A2)}$$
$$\text{Perceived}_{(\text{Bad},A1,A2)} = \text{Bias}_{(A1,A2,\text{Bad})} * \text{Designed}_{(\text{Bad},\ A2)}$$
$$\text{Perceived}_{(\text{Aid},A1,A2)} = \text{Bias}_{(A1,A2,\text{Aid})} * \text{Designed}_{(\text{Aid},\ A2)}$$
$$\text{Perceived}_{(\text{Obst},A1,A2)} = \text{Bias}_{(A1,A2,\text{Obst})} * \text{Designed}_{(\text{Obst},\ A2)}$$

To relate affordances to goals, the agent has goals it wants to achieve: desired goals. For example, an agent wants maintenance, power supply, and a knowledge base. The agent also has goals it wants to avoid: undesired goals. For example, a agent does not want to hurt its owner nor does it want to be destroyed.

### 6.2. Appraisal variables

Our agents look at other agents and users in terms of goal achievement. Agents have beliefs about goal-states in the world being true [0, 1] and about states facilitating or inhibiting other states [−1, 1] (−1: strong inhibition, 1: strong facilitation, 0: neutral). The *likelihood* of accomplishing a goal-state [−1, 1] (−1: Inhibit a goal state, 0: mean an event would not contribute to a goal state, but does not necessarily inhibit it, 1: facilitate a goal state) via another agent is calculated from the already-accomplished sub goals, which are states that facilitate or inhibit the goal-state, increasing or decreasing the likelihood of accomplishing the goal-state, respectively. All beliefs about the states being true are multiplied with the beliefs about the states facilitating or inhibiting the goal-state, according to algorithm A:

1. Sort the values in two lists: facilitation[0 → 1] and inhibition [0 → −1].
2. For both lists, start at 0 and take the mean of the value you have and the next value in the list until EOF.
3. Likelihood = weighed mean of the outcomes of both lists, with proportional weights (#pos/#tot) for the list of positive values, and (#neg/#tot) for the list of negative values.

Each facilitating sub goal that is achieved increases the perceived likelihood of reaching the goal-state, but the more sub goals are achieved, the less impact each has on the perceived likelihood (cf. Allen, 1934).

Agents also have beliefs that *actions of others* affect world-states [−1, 1] (−1: strong inhibition, 1: strong facilitation, 0: neutral). If an agent observes someone performing an action and believes this facilitates a certain world-state, the agent changes its *beliefs that the agent performing the action is responsible for establishing that world-state*. This is done according to formula F (see also Pontier & Siddiqui, 2009):

IF $\text{obs}_{(A1,A2,\text{Performs},\text{Action})}$
AND $\text{belief}_{(\text{Action},\text{Facilitates},\text{Goal-State})} > 0$
→ $\text{belief}_{(A2,\text{Responsible},\text{Goal-State})k+1} = \text{belief}_{(A2,\text{Responsible},\text{Goal-state})k} + \text{mf}_{\text{bel\_resp}} * \text{belief}_{(\text{Action},\text{Facilitates},\text{Goal-State})} * (1 - \text{belief}_{(A2,\text{Responsible},\text{Goal-State})k})$.

In formulas of the form F, $\text{mf}_{<\text{variable}>}$ [0, 1] is a *modification factor* that determines how quickly the variable is updated (here, believed responsibility). This modification factor is multiplied with the *impact value*, which is in factbelief $_{(\text{action},\text{facilitates},\text{goal})}$. Multiplying with *limiter* (1-old_belief) regulates going out of range, keeping an agent's beliefs from extremely high or low values.

Agents have beliefs about agents being *praiseworthy* [−1, 1], where −1 is blameworthy, 0 is neutral, and 1 is praiseworthy. If Coppélia believes a goal should have been achieved that in fact has not, she blames or praises Franz or her father who she believes are responsible for (not) achieving the goal. This is done according to formula F with as *impact value* belief $_{(A1,A2,\text{Responsible},\text{Goal-State})}$ * ambition_level$_{(\text{Goal-State})}$. If Franz (A1) beliefs that Dr. Coppélius (A2) is responsible for facilitating a desired goal-state or for preventing an undesired goal-state, Franz will increase his perceived praiseworthiness of Coppélius. In the reverse case, Franz will decrease his perceived praiseworthiness of Coppélius (and thereby increase his perceived blameworthiness).

### 6.3. Calculating expected utilities

In the system, the agents can perform actions to reach their goals. The system contains a library of goals, and each agent has a level of ambition for each goal. There are goals the agent wants to reach, and goals that the agent wants to avoid, all with several levels of importance. The levels of ambition the agent attaches to the goals are represented by a real value between $[-1, 1]$, where a negative value means that the goal is undesired and a positive value means that the goal is desired. A bigger value means the goal is more important for the agent.

The agents can perform actions to reach their goals. The system contains a library of actions from which the agents can choose. The agent has a belief about each action that it will inhibit or facilitate a certain goal. Its estimation of the facilitation of the goal by the action is represented by a real value between $[-1, 1]$, $-1$ being full inhibition, 1 being full facilitation. The following formulas are used to calculate the expected utilities of actions and features.

$$\text{ExpectedUtility}_{(\text{Action, Goal})} = \text{Belief}_{(\text{facilitates(Action, Goal)})} * \text{Ambition}_{(\text{Goal})}$$
$$\text{ExpectedUtility}_{(\text{Feature, Goal})} = \text{Belief}_{(\text{facilitates(Feature, Goal)})} * \text{Ambition}_{(\text{Goal})}$$

The expected utilities of features or possible actions are calculated by looking at the goal-states it influences. If an action or a feature is believed to facilitate a desired goal or inhibits an undesired goal, this will increase its expected utility. If it inhibits a desired goal, or facilitates an undesired goal, this will decrease its expected utility.

Given the level of ambition for a goal and the believed facilitation of a goal by an action towards another agent, the agent calculates the expected utility of performing that action towards that agent regarding that goal by multiplying the believed facilitation of the goal with the level of ambition for the goal.

A feature or action can have multiple expected utilities to several goals. Algorithm A calculates from all expected utilities of a feature the general expected utility as generalized across goals that are believed to be impacted by that feature. The general expected utilities of actions generate action tendencies in the agent with the same value.

Instead of discrete classification as positive approach, negative approach, change, or avoidance (Hoorn et al., 2008), actions have a continuous level of positivity and negativity. This allows for differentiating between constructive critique (positive change) and quibbling (negative change). To calculate general positivity and general negativity in the action tendencies, all action tendencies are multiplied with the positivity of the action and the negativity of the action in two separate lists. Algorithm A is performed to both these lists to calculate the general positivity and negativity of the action tendencies of the agent.

### 6.4. Similarity

For an agent to perceive its similarity with another agent, it needs to perceive the features of the self. Agents perceive their own features the same way they perceive the aesthetics and epistemics of other agents. Only this time, the bias is the bias in self-perception, instead of in the perception of another agent.

$$\text{Perceived}_{(\text{Feature,A1,A1})} = \text{Bias}_{(\text{A1,A1,Feature})} * \text{Designed}_{(\text{Feature, A1})}$$

Similarity is perceived according to the differences between the agent's perception of its own features and its perception of the features of the other agent:

$$\text{Similarity}_{(\text{A1,A2})} = 1\text{-}(\Sigma(\beta_{\text{sim}\leftarrow\text{feat}} * \text{abs}(\text{Perceived}_{(\text{Feat,A1,A2})} - \text{Perceived}_{(\text{Feat,A1,A1})}))$$
$$\text{Dissimilarity}_{(\text{A1,A2})} = \Sigma\ (\beta_{\text{ds}\leftarrow\text{feat}} * \text{abs}\ (\text{Perceived}_{(\text{Feat,A1,A2})} - \text{Perceived}_{(\text{Feat,A1,A1})}))$$

To calculate the dissimilarity between two agents, the differences between the perceived values for its own features, and those perceived for the other agent are taken. These differences are all added, with a certain (regression) weight β. Similarity is calculated in a similar manner, but with different weights, and the sum of all differences is subtracted of 1.

### 6.5. Relevance, valence, involvement and distance

The formulas in this paragraph were designed using generalized linear models (McCullagh & Nelder, 1989; Nelder & Wedderburn, 1972). Hoorn et al. (2008) shows that the calculated dependent variable (e.g., relevance) is fed by a number of contributing variables. Each contributing variable has a certain main effect on the dependent variable. The size of this main effect is represented by a (regression) weight $\beta$ the same way as for calculating similarity. When two variables interact, the interaction effect on the dependent variable is calculated by multiplying the product of the values of these two variables with a certain regression weight, which accounts for the interaction effect on the dependent variable. When the interaction is over-additive, the weight will be positive, and when it is under-additive, the weight will be negative.

The formula for the calculation of a variable A that is dependent on the variables B, C, and D, of which C and D interact, would be: $A = \beta_B * B + \beta_C * C + \beta_D * D + \beta_{CD} * C * D$. In this formula, $\beta_B$, $\beta_C$, and $\beta_D$ are the (regression) weights for the main effect of variables B, C, and D respectively, and $\beta_{CD}$ is the (regression) weight for the interaction effect of C and D. Hereby, $\beta_C$ is only that contribution from C that is controlled for D, and $\beta_D$ is the contribution of D that is controlled for C, and $\beta_{CD}$ is the contribution from C and D as an interaction effect. All chosen values for the $\beta$ weights are based on the theories the model is based on, such as the found effect sizes in the empirical studies leading to I-PEFiC.

In Hoorn et al. (2008), the action tendencies of each class have a direct effect on *involvement* (Inv) and *distance* (Dis), whereas here, the general positivity and negativity of action tendencies have an effect on involvement and distance via *relevance* (Rel) and *valence* (Val). If Coppélia has high expected utility for launching negative approach actions at Franz (beat him), this results in high general negativity of the action tendencies. This increases the negative valence that Coppélia experiences towards Franz, leading to an increase in distance and a decrease in involvement.

General action tendencies (GAT) are in the range [−1, 1] but are transformed to the range [0, 1] in these formulas by adding 1 and dividing the result by 2. All effects of variables on each other are summarized in Table 1.

Algorithm A computes the use intentions that, for example, Coppélia has with Franz, using the expected utilities of all Franz' features and the actions that she can perform to him. If Franz facilitates desired or inhibits undesired goal-states he will raise positive use intentions in Coppélia, and vice versa.

Table 1
The effects of variables on each other. In this table (+) stands for positive, and (−) for negative. DSim stands for Dissimilarity.

| | Rel | Irr | (+)Val | (−)Val | Inv | Dis |
|---|---|---|---|---|---|---|
| Beauty | | | | | × | × |
| Ugly | | | | | × | × |
| Realistic | | | | | × | × |
| Unrealistic | | | | | × | × |
| Good | × | × | × | × | | |
| Bad | × | × | × | × | | |
| (+)GAT | × | × | × | × | | |
| (−)GAT | × | × | × | × | | |
| Aid | | | | | × | × |
| Obstacle | | | | | × | × |
| (+)Val | | | | | × | × |
| (−)Val | | | | | × | × |
| (+)Val * Sim | | | | | × | × |
| (−)Val * Sim | | | | | × | × |
| (+)Val * DSim | | | | | × | × |
| (−)Val * DSim | | | | | × | × |
| (+)Val * Beauty | | | | | × | × |
| (−)Val * Beauty | | | | | × | × |
| (+)Val * Ugly | | | | | × | × |
| (−)Val * Ugly | | | | | × | × |
| Rel | | | | | × | × |
| Irr | | | | | × | × |
| Rel * Sim | | | | | × | × |
| Irr * Sim | | | | | × | × |
| Rel * DSim | | | | | × | × |
| Irr * DSim | | | | | × | × |
| Rel * Beauty | | | | | × | × |
| Irr * Beauty | | | | | × | × |
| Rel * Ugly | | | | | × | × |
| Irr * Ugly | | | | | × | × |

Her expected satisfaction is calculated by trading involvement (I) for distance (D), and taking a weighed mean of the involvement-distance trade-off (IDT) and the use intentions (UI) as described in Hoorn et al. (2008). When she does this for multiple agents, the agent with the highest expected satisfaction will be selected. Once an agent is selected, an action is searched, using:

$$\text{ExpectedSatisfaction}_{(\text{A1},\text{Action},\text{A2})} = w_{eu} * \text{Action\_Tendency} + w_{pos} * (1 - \text{abs}(\text{positivity} - \text{bias}_I * \text{Involvement})) + w_{neg} * (1 - \text{abs}(\text{negativity} - \text{bias}_D * \text{Distance}))$$

While looking for the strongest action tendency, the agent's positivity level seeks to come close to the level of (biased) involvement, the negativity level to (biased) distance. Shifts in weight can adjust the importance of positivity, negativity, and expected utility for selecting an action: The biases account for individual defaults (being a positively or negatively oriented person), which is a type of response modulation.

### 6.6. Effects on emotions

Believed likelihood that (un)desired world-states will occur underlie *hope* and *fear*. For all world-states with a believed likelihood, hope to achieve goals is calculated as:

$$\text{IF } f \geqslant \text{likelihood} \rightarrow \text{hope\_for\_goal} = -0.25 * (\cos(1/f * \pi * \text{likelihood}_{(\text{goal})}) - 1.5) * \text{ambition}_{(\text{goal})}$$

$$\text{IF } f < \text{likelihood} \rightarrow \text{hope\_for\_goal} = -0.25 * (\cos(1/(1 - f) * \pi * (1 - \text{likelihood}_{(\text{goal})})) - 1.5) * \text{ambition}_{(\text{goal})}$$

Here, $f$ is a shaping parameter (0, 1) that positions the top of the hope curve. The value of $f$ may differ per individual, representing 'fatalism' or 'pessimism': The top of the likelihood/hope-curve is always where likelihood $= f$. Thus, for $f$ close to 1, the top is situated at the right (representing hope only in cases of high probability); for $f$ close to 0, the top is left (representing hope even in cases of low probability). In our simulations, $f$ was set at 0.5. We chose a smooth instead of a linear function, because this matches human emotion curves (Bosse & Zwanenburg, 2009). Furthermore, a higher ambition simply leads to higher hopes (standard in the literature). Algorithm A is performed to the resulting values for hope_for_goal. Instead of step 3, however, hope is the outcome of the positive values list and fear is the absolute outcome of the negative values list.

If a world-state becomes true or false, the levels of *joy* and *distress* are calculated by formula F with ambition_level$_{(\text{World-State})}$ or − ambition_level$_{(\text{World-State})}$ as *impact value*. For instance, if a world-state becomes true, ambition_level$_{(\text{World-State})}$ determines *joy* and −ambition_level$_{(\text{World-State})}$ *distress*. A desired world-state that becomes true increases joy and decreases distress. An undesired world-state becoming true does the reverse.

This rule is applied also to world-states that facilitate other world-states, with $\text{belief}_{(\text{State,Facilitates,Goal-State})} * \text{ambition\_level}_{(\text{Goal-State})}$ or a negation of this multiplication as *impact value*. This way, achieving sub goals of desired world-states increases joy and decreases distress. Obstructive sub goals (world-states that inhibit goal-states) do the opposite. If a world-state becomes true, the agent's level of *surprise* moves towards the believed unlikelihood of this world-state happening, using:

$$\text{Surprise}_{K+1} = p_{\text{Surp}} * \text{Surprise}_k + (1 - p_{\text{surp}}) * (1 - \text{likelihood})$$

Here, $p_{\text{surp}}$ is a persistency factor, determining the slowness of adjustment of surprise. If an agent beliefs that a goal-state should have been reached, but it has not, this will increase its surprise according to F with $\text{likelihood}_{(\text{Goal-State})}$ as *impact value*. For all agents, the decay of surprise was set at 0.95 at each time step.

F also determined the level of *anger* [0, 1] with other agents, with $\text{belief}_{(\text{A2,Responsible,Goal-State})} * \text{ambition\_level}_{(\text{A1,Goal-State})}$ as *impact value*. If goal-states should have been achieved already, the agent gets angrier at those who are believed to be responsible and less angry at those who were helpful. If goal-states are undesired, the reverse happens. The decay factor of anger was set at 0.95 at each time step.

Algorithm A calculates the general level of anger from all levels of anger of one agent to another. Because there is only a list of positive values, step 3 becomes superfluous and the general level of anger simply is the outcome of step 2. Anger at self determines the value of *guilt*.

In Bosse et al. (2007), all agents have a desired level of intensity for each type of emotion. This is usually high for positive emotions (joy, hope) and low for negative emotions (anger, guilt). The overall mood is calculated by taking a weighed sum of the differences between the desired level and the actual level of emotion for all emotions simulated and subtracting this from 1:

$$\text{Mood} = 1 - (\Sigma(\beta_{\text{Emotion}} * \text{abs}(\text{actual}_{(\text{Emotion})} - \text{desired}_{(\text{Emotion})})$$

## 6.7. Emotion regulation strategies

Agents can perform *situation selection* and *situation modification* by affectively selecting situations and sub situations with the highest expected satisfaction. *Attentional deployment* shifts the focus of attention. Agents have beliefs that certain features cause emotions. If Coppélia focuses on Franz' dancing skills and her joy increases, this will increase her belief that Franz' dancing skills caused that emotion, using F with (Emotion(t) – Emotion(t – 1)) * Attention(Feature) as *impact value*. Emotion(t), then, is a level of emotion at a certain time point. Using the belief that a feature Feat causes an emotion E, an agent can shift attention Att as an emotion regulation strategy using:

$$\text{Att}_{(\text{Feat})k+1} = \text{Att}_{(\text{Feat})k} - \text{belief}_{(\text{Feat,causes,E})} * (\text{E} - \text{desired}_{(\text{E})})$$

This rule regulates that if Coppélia believes that Franz causes an emotion, she will pay more attention to him if she wants to increase that emotion, and less attention if she wants to down-regulate that emotion.

At each time step, relevance of features can also cause attention shifts by taking the absolute value of the general expected utility of a feature:

$$\text{Att}_{(\text{Feat})k+1} = p_{\text{att}} * \text{Att}_{(\text{Feat})k} + (1 - p_{\text{att}}) * \text{Relevance}_{(\text{Feat})}$$

Here, $p_{\text{att}}$ is a persistency factor, predicating the slowness of adjusting the attention. At each time step, the sum of the levels of attention is normalized to 1.

Cognitive change is implicitly performed by changing beliefs during the simulation (e.g., beliefs that actions facilitate goal-states, beliefs about the likelihood of goal-states, the praiseworthiness of others). These belief changes indirectly influence the agents' mood. Cognitive change can also be performed explicitly by changing the causal interpretation of past events, a form of emotion-focused coping. Suppose Franz feels guilty for not achieving the desired goal-state of being loyal (i.e., the level of guilt is above threshold), either because he performed an action that inhibited loyalty (court Swanilde), or did not perform an action that facilitated it (walk away from Swanilde). Then he can decrease his belief that an action (e.g., sleeping with Swanilde) had an influence on achieving the goal of being loyal by multiplying it with a modification factor, which is set at 0.8 for the current simulations.

### 6.7.1. Simulations

We implemented Silicon Coppélia in Javascript and performed simulation experiments under different parameter settings. We expected that using this model, a wider variety of behavior could be simulated than each of the three models it is based on. She is the concoction of sinister Doctor Coppélius, who made her so human-like that a young man named Franz was prepared to denounce his fiancée Swanilda for her.

Each experiment concerned the mechanic doll Coppélia, her creator Dr. Coppélius, and a young man, Franz, that is in a love affair with the doll. Coppélia and Franz are in a room. Franz is deliberating whether to allow or forbid Coppelia from going to a party. The possible world-states (i.e., the goals and subgoals the agents could have during the simulation) were 'Coppélia is having fun', 'Coppélia is safe', and 'Franz is happy'. The scenario was that there is a dance party going on and the possible actions that were inserted in the system were: Going to the party, allowing someone to go to the party, allowing someone to go to the party with some restrictions (e.g., be home early and not get kissed), and denying someone to go to the party. The results of the experiments are described below. Note that we sometimes used unrealistic settings to show how the model works.

*6.7.2. Baseline condition*

An initial experiment was performed that served as a control condition for the remaining experiments. In this condition, all parameters were set to 0, and the biases in perceiving features were set to the neutral value of 1. The desired levels of emotion were set to 0.8 for hope, 0.3 for fear, 0.9 for joy, 0.1 for distress, 0.4 for surprise, 0.1 for anger and 0.1 for guilt for all agents. The positivity and negativity of actions were defined according to Table 2.

The complete parameter settings for the baseline condition can be found in (Pontier & Siddiqui, 2009). This led to all emotions, perceived feature values, beliefs, expected utilities, action tendencies and general positivity and negativity in the action tendencies being 0. Because all the agents were exactly the same, the perceived similarity was 1 and dissimilarity was 0 for all agents. For all agents, the perceived relevance was 0.73, irrelevance was 0.28, and positive valence as well as negative valence was 0.29. The perceived involvement was 0.20, and the perceived distance was 0.12, leading to an involvement-distance tradeoff of 0.18. All use intentions were 0, together with the involvement-distance tradeoff leading to an expected satisfaction of 0.24 for all agents. All expected satisfactions for going to the party with another agent were 0.38, while all the expected satisfactions for the other actions the agents could perform were 0.40. The resulting mood level for all the agents was 0.62.

*6.7.2.1. Experiment 1: Franz gets angry.* In this experiment, Franz observed that Dr. Coppélius allowed Coppélia to go to a nightly dance party. Franz wanted his prospective bride to be safe (ambition level set to 1) and believed that allowing Coppélia to go to the party strongly inhibited this goal (belief set to −1). The results of this experiment are shown in Fig. 6.

Franz' belief led to a negative expected utility for allowing Coppélia to go to the party with respect to the goal of safeguarding his loved one, which led to a negative action tendency for this action. This resulted in Franz having negative use intentions towards Coppélia.

Compared to the baseline experiment, this decreased Franz' expected satisfaction of performing an action towards Coppélia, and the expected satisfaction of allowing Coppélia to go to the party.

Table 2
Positivity and negativity of actions.

| Action | Positivity | Negativity |
|---|---|---|
| Allow to go to the party | 0.8 | 0.2 |
| Allow to go with restrictions | 0.6 | 0.4 |
| Deny to go to the party | 0.2 | 0.8 |
| Go to party | 0.9 | 0.1 |

Because Franz observed that Dr. Coppélius allowed his creation to go to the party at such a late hour and because Franz believed that this inhibited the goal of his girlfriend being safe, Franz held Dr. Coppélius responsible for Coppélia not being safe. Because Franz wanted Coppélia to be safe, he thought Dr. Coppélius was to blame and he increased his bias of perceiving the badness of Dr. Coppélius and decreased the bias of perceiving the goodness of the inventor. (However, because the designed values for good and bad were set to 0 for all agents, the perceived goodness and badness did not change in this experiment). Franz got angry with Dr. Coppélius. Because of this, his general anger level increased. Thereby, his anger level moved further away from his desired level of anger, which causes a decrease in mood.

*6.7.2.2. Experiment 2: Coppélia's belief that states lead to other states.* In this experiment, dancing doll Coppélia wanted Franz to be happy (ambition level set to 1). She thought that if she were safe and was having fun, this would make Franz feel happy (both states were made sub goals of the state 'Franz is happy' with value 1). Due to the external event at time point 1, a music band marching into the street, Coppélia was having fun. At time point 2, Coppélia came home safely, which resulted in Franz being happy at time point 3. The results of this experiment are shown in Fig. 7.

At time point 1, because she was having fun, Coppélia believed that Franz might become happy. Because of this, she had hope for Franz becoming happy, which led to an increased level of general hope. Also, because Coppélia was having fun and none of the agents had any expectations that this would happen, their level of surprise increased. This increased the mood of Franz and Coppélia.

At time point 2, because she was having fun and was safe, Coppélia believed that Franz might become happy with a likelihood of 0.75. Because of this higher likelihood, she was pretty confident that Franz would become happy and therefore, her hope for Franz to become happy decreased, which also caused her general level of hope to decrease. Also, because Coppélia was safe and none of the agents had any expectations that this would happen, their level of surprise increased. This led to an increase of mood for Franz. Because Coppélia's hope decreased, this slightly decreased her mood.

At time point 3, Franz was even more surprised because he was happy. Coppélia, however, was already expecting him to become happy, so her level of surprise decreased. Because being happy was a desired goal of Coppélia, her level of joy increased, which elevated her mood.

*6.7.2.3. Experiment 3: Franz' involvement with Coppélia overrides his rationality.* In this experiment, Coppélia was a good, beautiful, realistic agent (designed features set to 1). Franz wanted her to be safe (ambition level set to 1) and

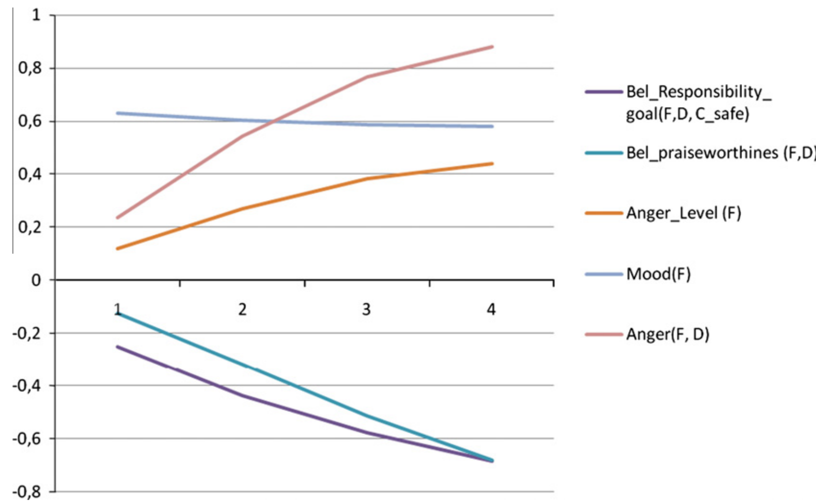



Fig. 6. Time plot of changing variables in experiment 1. In the time plots of the experiments, F = Franz, D = Coppélius, C = Coppélia).

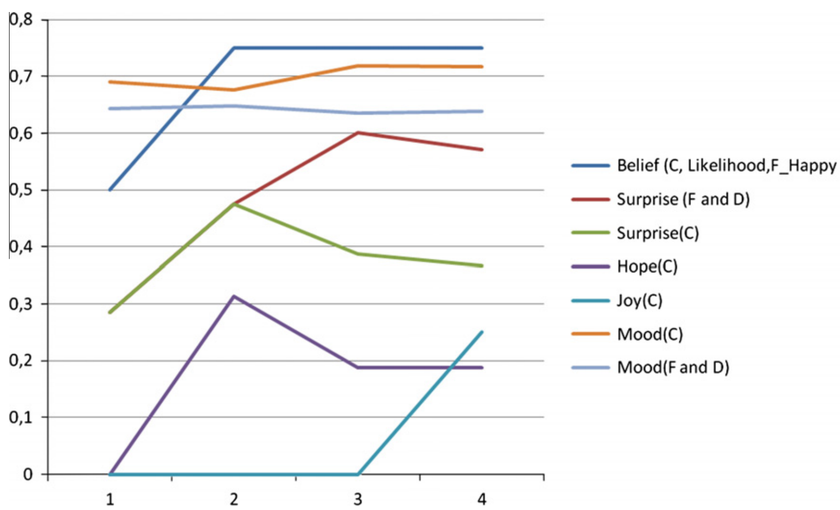



Fig. 7. Time plot of changing variables in experiment 2.

assumed that denying her to go to the late-night party facilitated this goal (belief set to 1). This led Franz to having an expected utility of 1 for forbidding Coppélia to go to the party with respect to the goal of her being safe and an action tendency of 0.5 for this action. Because of this, the general positivity and negativity in his action tendencies were respectively 0.03 and 0.10, and his Use Intentions were 0.25. The results of this experiment are shown in Fig. 8.

Due to the changed designed features compared to the baseline condition, the perceived similarity and relevance of Franz and Coppélia for each other decreased. Similarly, both the positive and the negative valence of Franz towards Coppélia increased.

This led Coppélia to having a decreased involvement towards Franz. The involvement and distance of Franz towards Coppélia both increased. Because of this, the output value of the involvement-distance trade-off of Coppélia towards Franz decreased, whereas Franz increased his value of the involvement-distance trade-off towards Coppélia.

The expected satisfaction of Franz for performing an action towards Coppélia increased. Whereas the expected satisfaction of denying Coppélia to go to the party decreased for Dr. Coppélius; for Franz, the expected satisfaction of performing this action increased, because of the high expected utility of this action. However, due to the increase in involvement, the expected satisfaction of allowing Coppélia to go to the party with restrictions increased to an even higher level. Therefore, Franz ended up allowing Coppélia to go to the midnight dance party with restrictions, where rationally he would have chosen to forbid her to go, because that action had a much higher expected utility for Franz, than the action he actually performed.

*6.7.2.4. Experiment 4: Coppélius and Franz disagree.* In this experiment, Coppélia was designed to be a beautiful agent (designed value set to 1). Compared to the baseline

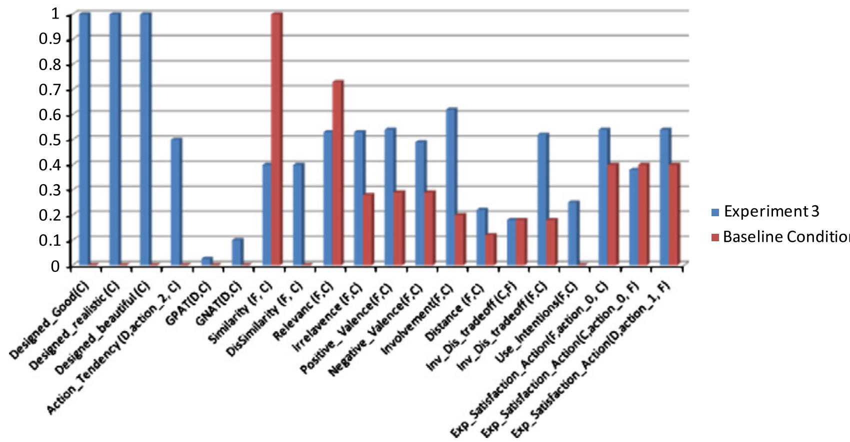



Fig. 8. Changes in variables in Experiment 3 compared to the baseline experiment.

experiment, this led both Franz and Coppélius to increase their attention to the beauty of their loved one, and therefore slightly decrease their attention to other features in the world. The results of this experiment are shown in Fig. 9.

Franz and Coppélius wanted Coppélia to have fun and to be safe (ambition level set to 1). Dr. Coppélius believed that his creation being beautiful facilitated the goal of her having fun (value set to 1). Franz, however, believed that this inhibited the goal of his girlfriend being safe (value set to -1). This led Franz to perceive a negative expected utility for the beauty of his lover regarding her being safe, which also caused a negative general expected utility for her beauty. Dr. Coppélius, however, perceived a positive expected utility for the beauty of his creation regarding the goal of Coppélia having fun, also causing a positive general expected utility for her beauty.

Further, Franz believed that allowing his girl to go to the party inhibited the goal of Coppélia being safe (value set to $-1$), and forbidding her to go to the party facilitated his girlfriend being safe. Coppélius believed, however, that allowing Coppélia to go to the party facilitated the goal of her having fun (value set to 1), whereas forbidding Coppélia to go to the party inhibited this goal (value set to $-1$).

This led Franz to have a negative expected utility for allowing Coppélia to go to the party regarding her safety, and he generated a negative action tendency for this action. He had a positive expected utility for forbidding his loved one to go to the party in view of her safety, and for this action he generated a positive action tendency.

For Dr. Coppélius, however, this led to a positive expected utility for allowing Coppélia to go to the party regarding her having fun, and generated a positive action tendency for this action. It also generated a negative expected utility for forbidding Coppélia to go to the party regarding her having fun, and he generated a negative action tendency for this action. This led both men to have a general positivity in action tendencies of 0.125, and a general negativity in action tendencies of $-0.125$.

Because Franz and Coppélius perceived their treasure as more beautiful than themselves, compared to the baseline experiment, their perceived similarity with Coppélia, and the perceived dissimilarity increased. This also increased

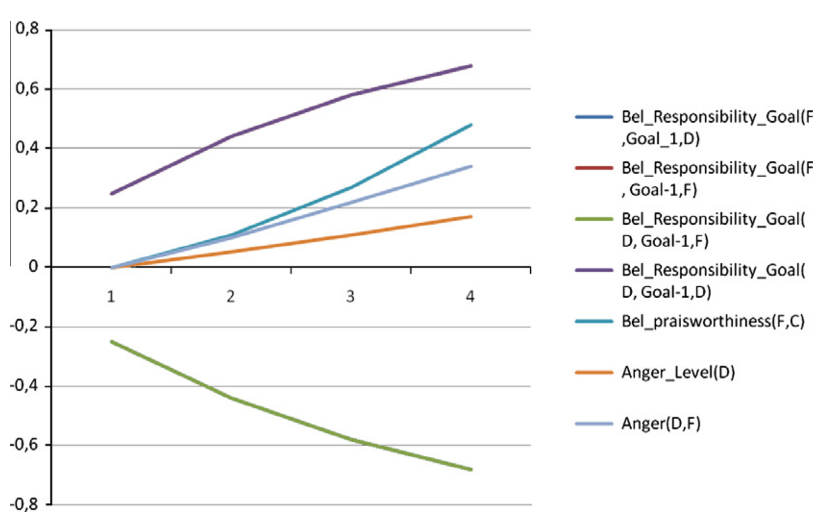



Fig. 9. Time plot of changing variables in Experiment 4. In the time plots of the experiments, Goal – 1 = Coppélia is safe).

their perceived positive valence towards Coppélia, and decreased their perceived negative valence towards her. This increased their involvement with their loved one, and decreased their distance. This led both men to have an increased involvement-distance tradeoff towards Coppélia. Franz had decreased intentions to interact with Coppélia, while Coppélius had increased use intentions. This led Franz to have a smaller expected satisfaction than Coppélius for interacting with Coppélia.

Compared to the baseline experiment, Franz decreased his expected satisfaction of allowing Coppélia to go to the party, and increased his expected satisfaction of forbidding Coppélia to go to the party. Dr. Coppélius increased his expected satisfaction of allowing Coppélia to go to the party, and decreased his expected satisfaction of forbidding her to go to the party. The expected satisfactions of the other two possible actions to perform to their loved one increased for both men.

This led Franz to forbid his girlfriend to go to the party, and Dr. Coppélius to allow his concoction to go. Therefore, Franz believed that Coppélius was responsible for inhibiting Coppélia's safety, whereas he held himself responsible for facilitating this goal. Dr. Coppélius, on his turn, held Franz responsible for inhibiting the goal of Coppélia having fun, whereas he held himself responsible for facilitating this goal.

Because both men kept on performing the same action, the next timestep these values were increased. Further, because Coppélius believed his creation should be having fun starting from this timepoint, while this was not the case, he blamed his would-be son-in-law for keeping Coppélia from having fun, whereas he praised himself for trying to make her have fun. Because of this, he got a bit angry at Franz, causing his general level of anger to increase.

Further, because he focused relatively much of his attention to the beauty of his creation while increasing his level of anger, Dr. Coppélius increased the belief that beauty causes anger, whereas this belief increased remarkably less for the other features.

*6.7.2.5. Experiment 5: Coppélius regulates his emotions.* For this experiment, the same parameter settings as in Experiment 4 were used. This time, however, Dr. Coppélius believed that allowing Coppélia to go to the party with some restrictions facilitated all goals inserted into the system. The results of this experiment are shown in Fig. 10.

The belief that allowing Coppélia to the party with some restrictions facilitated all goals, lead Coppélius to have a positive expected utility for this action regarding both the goals of Coppélia having fun and being safe. Therefore, he generated a tendency of 0.75 for this action. This increased his general positivity in action tendencies, which increased his perceived relevance of Coppélia, and decreased his perceived irrelevance of her. Further, it increased his perceived positive valence towards Coppélia. Because of this, his involvement with Coppélia increased, and his distance towards her decreased. This caused an increase of involvement-distance tradeoff. His intentions to interact with his creation were increased. This led to an increased expected satisfaction for interacting with Coppélia.

Compared to Experiment 4, Dr. Coppélius increased his expected satisfaction of allowing Coppélia to go to the party, and performed this action. This caused him to belief he was responsible for facilitating all goals inserted into the system. Also, Franz did not see Coppélius anymore as inhibiting his girlfriend having fun, because Franz did not have any beliefs about allowing Coppélia to go to the party with some restrictions.

Because Dr. Coppélius kept performing the same action, the next timestep these values were increased. Further, Dr. Coppélius believed that Coppélia should be having fun, while this was not the case. Because he did not allow her

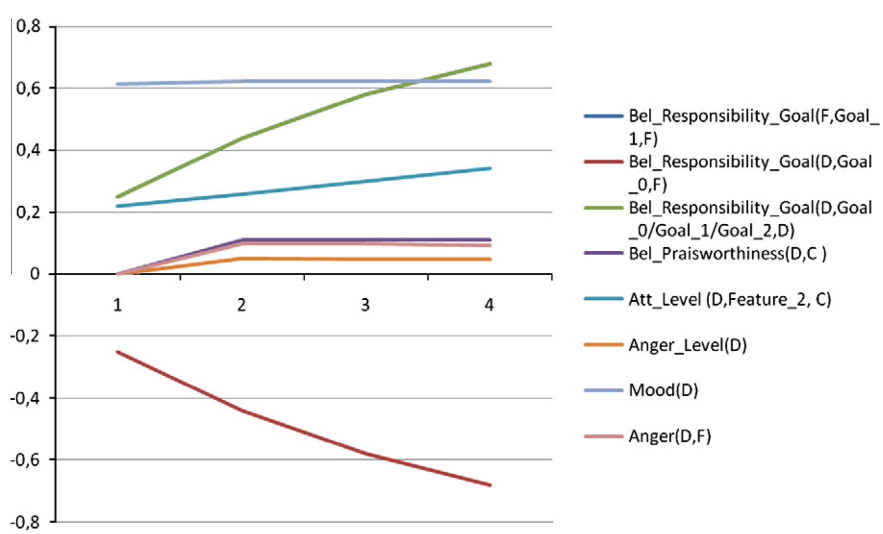



Fig. 10. Time plot of changing variables in Experiment 5. In this time plot, Goal_0 = Coppélia is having fun, Goal_2 = Franz is happy, and Feature_2 = beauty).

to go to the party without restrictions while he believed this would facilitate Coppélia having fun, Dr. Coppélius regulated his emotions. As an emotion-regulation strategy, he decreased the belief ‘allowing Coppélia to go to the party without restrictions will lead her to having fun’.

## 7. Conclusion

We presented an implementation of Silicon Coppélia, which is a combination of CoMERG (Bosse et al., 2007), I-PEFiC$^{ADM}$ (Hoorn et al., 2008), and a simplified version of EMA (Gratch & Marsella, 2006; Marsella & Gratch, 2009). In this model, the agents have goal-related beliefs that lead to emotions. More specifically, the agents have beliefs about the responsibility of other agents for (not) achieving goal-states, and about praiseworthiness based on this responsibility. The agents also have beliefs about the likelihood of achieving goal-states. These beliefs were based on other beliefs that were also inserted into the system. The beliefs together with the other variables in the system affected the seven emotions that were experienced by the agents: joy, distress, hope, fear, anger, guilt and surprise. These emotions were aggregated into an overall mood. Further, certain emotion-regulation strategies were added to the system based on Bosse et al. (2007), and Gratch and Marsella (2006, 2009).

Simulation of the behavior of Silicon Coppélia showed that in Experiment 1, Franz became cross with Dr. Coppélius, because in his view, the inventor performed an action (allowing Coppélia to go to a dance party) that was in conflict with Franz’ goals (Coppélia being safe). In Experiment 2, changes in the world led Coppélia to have beliefs about the likelihood of other world-states becoming true. This caused her to experience hope, and later in the simulation joy when the expected and desired goal-state became true. Further, all the agents experienced surprise when world-states unexpectedly became true. In Experiment 3, due to involvement with Coppélia, Franz made the affective decision to allow his girlfriend to go to the midnight party, where rationally, he would have forbidden her to go. In Experiment 4, Franz and Dr. Coppélius disagreed on what was best for their loved one, and therefore Coppélius becomes a bit angry at Franz, as he believed That Franz kept Coppélia from having fun. In Experiment 5, Coppélius did not perform an action of which he believed it would help to reach his goals, because he saw a better option. As an emotion-regulation strategy, later in the simulation he decreased the belief that this action would have helped.

Previous experiments showed that CoMERG (Bosse et al., 2007), I-PEFiC$^{ADM}$ (Hoorn et al., 2008) and EMA (Gratch & Marsella, 2006; Marsella & Gratch, 2009) are not able to simulate this kind of behavior on their own. CoMERG and I-PEFiC$^{ADM}$ cannot simulate emotions or belief-changes based on beliefs about the responsibility of other agents and the likelihood of goal-states happening (Experiments 1–5). EMA is not capable of making irrational decisions when appropriate (Experiment 3). Therefore, we may conclude that the simulation experiments confirmed that Silicon Coppélia shows richer affective behavior than the other models.

We have conducted an initial study which indicates that an agent equipped with the combined model can behave emotionally believable (Pontier, Siddiqui, & Hoorn 2010).

However, one fallacy of our approach obviously is that ‘richer’ does not necessarily mean ‘better’. As often in computational modeling, all other things being equal, the simplest solution is the best. Thus, the combined model of affect should only contain those variables of which it is certain that they have an added value. User studies will have to point out whether this is the case. Therefore, we plan to perform systematic tests to assess whether in the eyes of the user Silicon Coppélia indeed results in more human-like affective behavior than the three sub models do separately.

As soon as the model has been validated in user studies, we will start exploring the possibilities to apply it to real humans instead of agents; i.e. to develop a robot that can communicate affectively with humans in a more natural way, with a mind of its own, in pursuit of its own goals, and acting emotionally intelligent.

## References


Allen, R. G. D. (1934). The nature of indifference curves. *The Review of Economic Studies, 1*(2), 110–121.

Beck, A. T. (1987). Cognitive models of depression. *Journal of Cognitive Psychotherapy, an International Quarterly, 1*, 5–37.

Bosse, T., & Zwanenburg, E. (2009). There’s always hope: Enhancing agent believability through expectation-based emotions. In: M. Pantic, A. Nijholt, J. Cohn (Eds.), *Proceedings of the 2009 international conference on affective computing and intelligent interaction, ACII’09* (pp. 111–118). IEEE Computer Society Press.

Bosse, T., Pontier, M., Siddiqui, G. F., & Treur, J. (2007). Incorporating emotion regulation into virtual stories. In: C. Pelachaud, J. C. Martin, E. Andre, G. Chollet, K. Karpouzis, & D. Pele (Eds.), *Proceedings of the seventh international conference on intelligentvirtual agents, IVA’07: Vol. 4722. Lecture notes in artificial intelligence* (pp. 339–347). Springer Verlag.

Bosse, T., Pontier, M., & Treur, J. (2007). A dynamical system modeling approach to Grosśmodel of emotion regulation. In: R. L. Lewis, T. A. Polk, & J. E. Laird (Eds.), *Proceedings of the 8th international conference on cognitive modeling, ICCM’07* (pp. 187–192). Taylor and Francis.

Ekman, P., & Rosenberg, E. L. (2005). What the Face Reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). New York: Oxford University Press.

Endsley, M. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*, 32–64.

Frijda, N. H. (1986). The emotions. New York: Cambridge University.

Frijda, N. H. (1988). The laws of emotion. *American Psychologist, 43*, 349–358.

Gratch, J., & Marsella, S. (2006). Evaluating a computational model of emotion. *Journal ofAutonomous Agents and Multiagent Systems (Special issue on the best of AAMAS (2004), 11*(1), 23–43.

Gross, J. J. (2001). Emotion regulation in adulthood: Timing is everything. *Curent Directions in Psychological Science, 10*(6), 214–219.

Hoorn, J. F., Pontier, M. A., & Siddiqui, G. F. (Eds.). (2008). *When the user is instrumental to robot goals. First try: Agent uses agent* Jain, L., Gini, M., Faltings, B. B., Terano, T., Zhang, C., Cercone, N., & Cao, L. (Eds.). *Proceedings of IEEE/WIC/ACM Web Intelligence and Intelligent Agent Technology 2008 (WI-IAT ’08)*

(Vol. 2. 978-0-7695-3496-1, pp. 296–301). Sydney AU: IEEE/WIC/ACM. doi:10.1109/WIIAT.2008.113.

Konijn, E. A., & Hoorn, J. F. (2005). Some like it bad. Testing a model for perceiving and experiencing fictional characters. *Media Psychology, 7*(2), 107–144.

Lazarus, R. S. (1991). Emotion and adaptation. New York: Oxford University.

Mao, W., & Gratch, J. (2006). Evaluating a computational model of social causality and responsibility. In: *5th international joint conference on autonomous agents and multiagent systems, Hakodate, Japan*.

Marsella, S., & Gratch, J. (2003). Modeling coping behavior in virtual humans: Don't worry be happy. In *Proceedings of Second International Joint Conference on Autonomous Agents and Multiagent Systems. AAMAS'03* (pp. 313–320). ACM Press.

Marsella, S., & Gratch, J. (2006). EMA: A computational model of appraisal dynamics. In *Agent Construction and Emotions*, Vienna, Austria.

Marsella, S., & Gratch, J. (2009). EMA: A model of emotional dynamics. *Cognitive Systems Research, 10*(1), 70–90.

McCullagh, Peter, & Nelder, John (1989). *Generalized Linear Models. 2nd ed.*. Boca Raton: Chapman and Hall, ISBN: 0-412-31760-5.

Nelder, John, & Wedderburn, Robert (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General), 135*(3), 370–384 (Blackwell Publishing).

Pontier, M. A., & Siddiqui, G. F. (2009). IAT-2009 Appendix. Retrieved from <http://www.few.vu.nl/~ghazanfa/iat2009_appendix.html>.

Pontier, M. A., Siddiqui, G. F., & Hoorn, J. (2010). Speed dating with an affective virtual agent – Developing a testbed for emotion models. In: J. Allbeck, et al. (Eds.), *Proceedings of the 10th international conference on intelligent virtual agents, IVA'10: Vol. 6356. Lecture notes in computer science* (pp. 91–103).

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. In P. Shaver (Ed.). *Review of Personality and Social Psychology* (Vol. 5, pp. 11–36). Beverly Hills, CA: Sage.

Scherer, K. R. (1993). Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition and Emotion, 7*, 325–355.

Smith, C. A., & Lazarus, R. S. (1990). Emotion and adaptation. In L. A. Pervin (Ed.), *Handbook of personality: Theory & research* (pp. 609–637). NY: Guilford Press.

Van Vugt, H. C., Hoorn, J. F., Konijn, E. A., & De Bie Dimitriadou, A. (2006). Affective affordances: Improving interface character engagement through interaction. *International Journal of Human-Computer Studies, 64*, 874–888.

Van Vugt, H. C., Hoorn, J. F., & Konijn, E. A. (2009). Interactive engagement with embodied agents: An empirically validated framework. *Computer Animation and Virtual Worlds, 20*, 195–204. doi:10.1002/cav.31.