An Interdependent Model of Personality, Motivation, Emotion, and Mood for Intelligent Virtual Agents

Maayan Shvo  
University of Toronto  
Toronto, Canada  
maayan.shvo@cs.toronto.edu

Jakob Buhmann  
DisneyResearch|Studios  
Zurich, Switzerland  
jakob.buhmann@disneyresearch.com

Mubbasir Kapadia  
Rutgers University  
Piscataway, New Jersey  
mubbasir.kapadia@rutgers.edu

ABSTRACT
Building intelligent agents that can believably interact with humans is a difficult yet important task in a host of applications, including therapy, education, and entertainment. We submit that in order to enhance believability, the agent’s affective state should be accurately modeled and should realistically influence the agent’s behavior. We propose a computational model of affect which incorporates an empirically-based interplay between various affective components - personality, motivation, emotion, and mood. Further, our model captures a number of salient mechanisms that are observable in humans and that influence the agent’s behavior. We are therefore hopeful that our model will facilitate more engaging and meaningful human-agent interactions. We evaluate our model and illustrate its efficacy, as well as the importance of the different components in the model and their interplay.

CCS CONCEPTS
• Human-centered computing → Human computer interaction (HCI).

KEYWORDS
Intelligent Virtual Agent; Model of Affect

1 INTRODUCTION
Decades-long research in artificial intelligence has sought to build believable, emotional, and engaging intelligent agents. Research on emotional agents has a myriad of applications. Si [34] uses emotionally complex agents employing theory of mind in a multi-agent framework for authoring and simulating interactive narratives. Aylett et al. [1] demonstrate how emotional agents can be used in an anti-bullying educational module. However, there exist many challenges when creating computational models of affect for intelligent agents [19] such as: (1) faithfully modelling human emotions in a way that is not foreign to a human; and (2) creating a believable affect-consequent model - a mapping of the agent’s affective state into behavioral and cognitive changes.

Throughout the years, multiple computational models of affect have been proposed that differ in the modeled affective components and the relations between them (e.g., [1, 8, 11–13]). However, due to insufficient modeling of the relations between various affective components in the model, previous work has struggled to capture a number of important mechanisms which influence behavior and that are evident in human beings.

We introduce in this work a computational model of affect which comprises various affective components - personality, motivation, emotion, and mood - and an empirically-based interplay between the different components. The personality model is based on the Five Factor Model (FFM) [21]; the motivational model is based on Reiss’ theory of 16 basic desires [31]; and the emotional model is based on appraisal theory and the PAD (Pleasure, Arousal and Dominance) theory [24], where emotion and mood represent different temporal resolutions of the agent’s affect. Crucially, the model captures a strong interplay between the aforementioned components, as shown in Figure 1. Personality influences the intensity of experienced emotions, biases the default mood of the agent, and modulates the importance of the various motivations; mood is primarily modulated by the agent’s past emotions and influences the intensity of currently experienced emotions; finally, intense emotions elicit emotional reactions and influence the importance given to the different motivations, which are the basis of the agents’ goal-driven decision making. The interplay between the various affective components allows for generalization across a myriad of emotional agents, differing in their disposition and motivations. While the implementations of the relations and components in our model can be replaced by alternative theories and formalizations, their inclusion in the model is crucial in that they allow the model to capture important mechanisms that influence the agent’s decision making and are observable in human beings [21, 22, 26, 27].

Our model includes an affect-consequent mechanism which is grounded in therapeutic psychological literature and which enables the agent to address the root cause of the negative emotion it is experiencing. This mechanism is realized by algorithmically bridging between the motivation-based planner, which drives the agent’s decision making, and the agent’s affective state. The heuristic-based forward planner generates agent behavior (a plan) which achieves its desired motivational goal and operates within a 16 dimensional continuous space defined by the agent’s motivations. The heuristic...
attempts to maximize the agent’s motivation values, given a weighting over the 16 dimensions. We allow the agent’s affective state to influence the planner by formalizing a relationship between each of the 16 motivations and an emotion type. Thus, when a certain emotion type is experienced with sufficient intensity, we modify the weighting given to the motivational dimensions which correspond to the emotion type. The modified weighting influences the search and causes it to prefer actions which more heavily influence the respective motivations.

Finally, we demonstrate the impact of the various components in the model and their interaction with one another on the agent’s decision making. We show that the agent displays disparate behavior with and without the inclusion of specific components in our model. The main contributions of this paper are as follows: (1) we propose a computational model of affect which incorporates a novel empirically-based interplay between the model’s various affective components, allowing for generalization across a myriad of emotionally complex agents; (2) we propose a novel affect-consequent model which is grounded in therapeutic psychological literature and facilitated by an algorithmic bridging between the agent’s affective and motivational states; (3) we weight the agent’s motivational profile with its FFM personality traits, using an empirically based correlation between the 16 Reiss desires and the FFM traits.

2 RELATED WORK
Numerous computational models of affect have been proposed by previous work. For example, the ALMA model [11] temporally differentiates between three layers of affect - personality, mood, and emotion - and models the relations between the three. Additionally, the WASABI model [2] differentiates between primary and secondary emotions and combines them with facial expressions. The influential EMA model [20] attempts to capture the complete range of human emotion, in addition to being the first computational model of affect to include emotional coping.

Additionally, previous work in the field has incorporated Reiss’ theory of basic desires into agent architectures; for example, in [30], the authors use the 16 basic desires to model the agent’s personality. However, in their work the relationship between personality traits and basic desires is not empirically based since the FFM theory is not used. Further, both in [35] and [6] the Reiss basic desires are used as the agent’s sole personality model.

Previous work has incorporated a subset of the affective components and the relations between them which can be found in our model; thus, it has struggled to capture a number of important mechanisms which influence behavior and that are observable in human beings. For example, while ALMA incorporates the influence personality has on an individual’s emotion and mood, it does not incorporate a motivational profile for the agent, nor does it include an affect-consequent model which draws a correspondence between the type of emotion experienced by the agent, and its motivations. While related to our work, Lim et al.’s approach [16] does not temporally differentiate between three layers of affect nor does it handle negative emotions in the novel way we describe here.

To address negative emotions, previous work mainly focuses on the implementation of various coping strategies found in psychological literature [3, 20]. These strategies chiefly attempt to mitigate negative emotions either by changing features in the environment that led to the initial undesirable appraisal (problem-focused) or by altering cognitive processes (e.g., wishful thinking, resignation, or suppression of existing negative emotions) (emotion-focused). To contrast between our affect-consequent model and previous approaches, consider the following example: an agent is experiencing shame and feeling deficient or inferior - the root cause of the emotion, as found by research [17]. The various coping strategies implemented by previous work might address this emotion by, e.g., suppressing the emotion or treating the supposed antecedent, namely trying to rectify or undo what has been done by the agent. However, when shame is suppressed it tends to worsen and undoing what has been done does not address the root cause of the emotion.

To address the feeling of inferiority or deficiency, the agent needs to feel accepted by its environment by, for example, seeking reassurance, leading to increased acceptance. Shame was also treated in Conati [7], where an agent offers hints to an ashamed student to prevent her from feeling future shame. Our approach instead treats the current shame experienced by the agent; in the future, our model could enable an agent to help another agent address the root cause of her shame. In general, our approach addresses the root cause of the emotion being experienced by the agent, by drawing a correspondence between its emotions and motivations. Note that previous work does not map the agent’s emotion types to Reiss’ 16 basic desires. Additionally, research suggests that while personality and motivation, specifically the FFM traits and Reiss’ basic desires, are highly correlated, they do not seem to map onto one another such that only one of the two can be used to model both [4, 27]. Therefore, there is merit to modeling both the agent’s personality and its motivations, while weighting the agent’s motivational profile according to the empirically based correlations found between the two constructs. Finally, we are able to model the empirically based influence personality has on the agent’s emotional state.

3 OVERVIEW
A generic intelligent agent architecture, as described by Cassell et al. [5], comprises a number of standard modules: (1) an input manager which monitors various sensory input from the environment; (2) a deliberative module which is tasked with the agent’s decision making and tracking the agent’s state; and (3) an action scheduler which is in charge of executing all agent behavior and interacting with the agent’s environment. The focus of our work, namely the interplay between personality, motivation, emotion, and mood, is modeled within the deliberative module and affects the agent’s decision making. For modules (1) and (3), we use a simple baseline implementation, and the integration of our model with a more complex architecture is left to future work.

Figure 1 presents the interplay between the various affective components and places our model within a generic agent architecture, which conforms to the architecture described above. We model the influence of personality on the emotional state of the agent based on research that has shown that individuals with different personality traits experience emotions differently [21]. Additionally, the model accounts for an impact of an agent’s personality...
on its motivation which reflects a correlation between an individual’s personality traits and the importance they assign to each of Reiss’ 16 basic desires [27]. Following empirical findings that mood affects the intensity of an experienced emotion [26], our model incorporates an interaction between mood and emotions. Further, the agent’s active emotions change its mood such that it faithfully represents this empirically-based definition of mood: “an average of a person’s emotional states across a representative variety of life situations” [22]. Lastly, emotions influence motivations via an affect-consequent model which is grounded in therapeutic psychological literature [17] and which formalizes a relationship between the agent’s behavior-driving motivations and its affective state.

4 AFFECTIVE COMPONENTS

In this section, we discuss each of the affective components in our model: Personality (P), Motivation (M), Emotion (E), and Mood (B). Note that the agent’s state, is defined by an ever-updating logical knowledge base and is a conjunction of propositions that are true in the world. The state representation is used by the planner for node expansion, however, that is not the focus of this work.

4.1 Personality

We model the agent’s personality, P, using the Five-Factor Model (FFM), which is a comprehensive classification of personality traits that is currently the most widely used dimensional model of personality [21]. The five personality traits are: Neuroticism, Extraversion, Conscientiousness, Agreeableness, and Openness. We represent this as a vector \( p \in P \), where each trait, \( p_i \), is assigned a value between [0, 1]. In the following sections, we will discuss how the agent’s personality profile influences its emotional state and behavior.

4.2 Motivation

A myriad of factors influences human decision making, specifically the way in which we choose goal-oriented actions. Chief among them, perhaps, are our intrinsic desires and motivations [31]. Note that we use the terms ‘desire’ and ‘motivation’ interchangeably throughout the paper. There exist many theories of motivation in the literature (e.g., [18, 32]). In this work, we model the agent’s underlying motivations using Reiss’ empirically tested theory of 16 basic desires [31]. Reiss’ theory was initially chosen as it was suitable for rapid prototyping of our system, as well as its integration within a logic-based system. The 16 desires are: Power, Curiosity, Independence, Status, Social contact, Vengeance, Honor, Idealism, Physical exercise, Romance, Family, Order, Eating, Acceptance, Tranquility, and Saving. It is posited by this theory that any behavior is executed to satiate one or more of these basic desires. To incorporate the idea that an agent strives to fulfill its basic desires and motivations and to have this notion drive the agent’s decision making, we model the agent’s motivational state using two 16-dimensional vectors. One of the vectors, \( m^c \in M \), represents the current level of motivation fulfillment and the second, \( m^d \in M \), represents the desired, or target, level of motivation fulfillment. Each dimension in these vectors is assigned a value between [0, 1], and the objective is to minimize the distance between \( m^d \) and \( m^c \). Changes in \( m^c \) are achieved by reaching specific states that encode (rule-based) motivational changes. E.g., eating will have a positive effect on the eating dimension in \( m^c \). Additionally, it has been shown that individuals behave differently because they assign different importance to each basic desire [31]. We model this notion of importance with a linear weighting of the 16 motivations when computing the distance between \( m^d \) and \( m^c \).

**Motivation-driven Planning**

We utilize a forward planner with an A’ heuristic search to generate a plan \( \pi_G \), where the heuristic leverages a weighted distance between \( m^c \) and \( m^d \):

\[
d(m^c, m^d, w) = \left( \sum_{i=1}^{16} w_i \cdot (m^d_i - m^c_i)^2 \right)^{1/2},
\]

where the weight vector, \( w \), defines the importance given to each of the dimensional values, as determined by the agent’s personality and emotional state (see Section 5). The distance between \( m^c \) and
\( \mathbf{m}^d \) represents how far the agent is from fulfilling its motivations, and the search algorithm attempts to choose actions that decrease this distance by reaching states which are mapped to certain motivation values. Once the heuristic search has found a plan that sufficiently satisfies the current motivations (w.r.t. the desired motivations), the plan is sent to the behavior manager, which uses a simple priority based system to prioritize behavior execution. E.g., if there is an ongoing plan and an emotional (or physical) reaction, as will be described in Section 4.3, is sent to the behavior manager, it will interrupt the plan in order to execute the emotional reaction.

### 4.3 Emotion

We assume that emotions reflect a short-term affect that arises as a consequence of life situations, i.e. stimuli from the environment. Thus, our appraisal mechanism, tasked with evaluating the state of the agent (given by a knowledge base) and subsequently generating relevant emotions, uses a set of predefined rules that map a state in which the agent might be, to an instance of an emotion type \( e \in E \) and its intensity \( I(e_i) \). The 21 emotion types \( e_i \) used in this work are taken from the OCC theory [28] and differentiated by \( i \). After an emotion \( e_i \) is generated at \( t_0 \) and assigned an initial intensity \( I_0(e_i) \), that intensity is decayed as a function of time \( t \) [29].

\[
I_t(e_i) = I_0(e_i) \cdot e^{-\beta(t-t_0)},
\]

where the constant \( \beta \) determines how fast the intensity of the emotion \( e_i \) will decrease over time. Once an emotion is generated, it is stored in the list of active emotions \( \epsilon \) until its intensity falls below a predefined threshold near zero. Finally, as will be described in Section 5.3, the initial intensity of an emotion, \( I_0(e_i) \), is modulated by the agent’s personality traits \( \mathbf{p} \) and current mood \( \mathbf{b} \).

#### Acting on Emotions

Linehan [17] posits that emotions lead to reactive action urges. E.g., extreme anger may lead to actions such as yelling, or even throwing an object, at another agent. Our model includes an emotionally reactive layer - situated within the deliberative module - to allow the agent the express its emotions. Based on the experienced emotion and the situational context of the agent, the reactive layer quickly selects an apt emotional reaction. For instance, if the agent is by itself and extremely angry, it might punch a wall in fury, however, when in the presence of another agent, it might yell at the fellow agent. The reactive layer monitors the intensity of the active emotions \( \epsilon \), and when the intensity of \( e_i \in \epsilon \) crosses an emotional reaction threshold \( \tau_e \) (s.t. \( I(e_i) > \tau_e \)), the reactive layer propagates a goal \( g_{e_i} = \text{(emotional-reaction } e_i \text{)} \) to a goal planner to produce an appropriate plan \( \pi_{\epsilon}(g_{e_i}) \). This is defined more formally in Algorithm 1. Compared to the heuristic-based planner, this planner does not operate in the 16-dimensional motivation space, but rather has a fixed goal in the agent’s state space and operates in a standard STRIPS style backward search [9]. The planner recursively satisfies all goal conditions of the emotional reaction while constantly including all unsatisfied preconditions of an expanded affordance as goal conditions. While the emotional reaction is currently executed immediately, the suppression or regulation of an emotional reaction is left to future work.

### 4.4 Mood

In this work, we compute mood similarly to the ALMA computational model [11]. Mood is seen as a medium-term affective state, building on the model proposed by Mehrabian [22, 25]. Mood is distinguished from emotion by its resolution and stability over time: while mood reflects the agent’s relatively stable and lasting affective state, the agent’s experienced emotions can often be volatile and ever-changing. Mehrabian describes mood with the three traits Pleasant (P), Arousal (A), and Dominance (D) that have been shown to be nearly independent [22].

We implement the mood \( \mathbf{b} \in B \) as a vector in this PAD space where each dimension ranges from -1 to 1. Additionally, we may refer to a discretized description of the current mood \( \mathbf{b} \) (see Table 1 in the Supplemental Material). Finally, following ALMA [11], we define an agent’s initial default mood, \( \mathbf{b}_0 \), based on a mapping between the FFM personality traits and the PAD mood space, as found in [23]. In Section 5.2, we will discuss how mood is dynamically computed based on the agent’s active emotions and personality.

### 5 THE INTERPLAY OF PERSONALITY, MOTIVATION, EMOTION, AND MOOD

In this section, we offer an in-depth look at the specifics of Figure 1, discussing the interplay between the various affective components. Algorithm 1 outlines the inner workings of the deliberative module, consisting of the various mechanisms described in this section and in the previous section. At every time step, the active emotions \( \epsilon \) are decayed and filtered, while newly generated emotions are added. Then, if a sufficiently intense emotion exists, an emotional reaction \( \pi_{\epsilon} \) could be planned for, or, alternatively, the motivation weight vector \( \mathbf{w} \) could be modified based on the experienced emotion. Finally, if an emotional reaction is not underway, the \( \mathcal{A}^* \) planner is tasked with finding a motivationally-driven plan \( \pi_{\epsilon} \). The resulting plan is then sent to the behavior manager and finally to the action scheduler for execution. Note that we do not focus in this work on prioritizing and scheduling different plans.

#### 5.1 Personality and Motivation

We base the influence of personality on motivation on an empirically-based correlation between a person’s personality traits and the importance they assign to each motivation [4, 27]. For instance, following the statistically significant findings in [27], we positively correlate extraversion with social contact, power, and status. Similar mappings exist for the other four personality traits (see Table 2 in the Supplemental Material for the full mapping). The weight vector \( \mathbf{w} \), that is used by the heuristic search, reflects this correlation. \( \mathbf{w} \) is used in Equation (1) to weight the individual motivational dimensions; therefore, when the planner tries to minimize that distance, the personality traits will have a direct impact on the resulting plan. E.g., an agent with high extraversion would weight the dimensions more strongly, which will cause the planner to generate plans that bring \( \mathbf{m}^d \) more closely to \( \mathbf{m}^p \) along those dimensions. \( \mathbf{w} \) is modified given the agent’s personality profile in step 2 of Algorithm 1.

---

\footnote{The Supplemental Material can be found in: https://zenodo.org/record/2672901# XNHPy98rgWp}
where which is used at each time step. To simulate human-like mood changes, we model the mood-similarity of a given emotion to the current mood:

\[
\text{sim}(\epsilon, m) = \frac{\text{Similarity}(\epsilon, m)}{\text{Max Similarity}}
\]

where \( \text{Similarity}(\epsilon, m) \) is a measure of similarity between the emotion \( \epsilon \) and the current mood \( m \). This similarity is calculated based on the PAD space, which reflects the joint set of all active emotions and their intensities. The PAD space is defined as:

\[
PAD = \{ \epsilon \mid \epsilon \in \text{Active Emotions} \}
\]

where each element \( \epsilon \) represents an emotion with its intensity and valence.

5.3 Mood and Personality as Emotional Modulators

As defined in Equation (2), the intensity of an active emotion exponentially decays over time from an initial value \( I_0 \). However, the initial strength can vary largely between the different emotions. This variation is captured based on the pre-authored emotional trigger which generates the emotion, the current mood, and the personality:

\[
I_k(e_i) = I_k \cdot \left( I_{b \rightarrow e_i} + I_{p \rightarrow e_i} \right) / 2,
\]

where \( I_k \) describes the pre-authored strength of the \( k \)th emotional trigger, \( I_{b \rightarrow e_i} \) and \( I_{p \rightarrow e_i} \) are multiplicative influences based on the agent’s mood and personality, respectively. While the emotional trigger holds the most influence over the emotion’s intensity, the influence of the agent’s mood and personality over the intensity offers an empirically-based way in which to incorporate differences across agents, as well as varying emotional context. To this end, we used the following formalism:

\[
I_{b \rightarrow e_i} = 1 + \|b_t\| \cdot \left( I(\phi(e_i), b_t) - I(\phi(e_i), \bar{b}_t) \right)
\]

\[
I_{p \rightarrow e_i} = 1 + \sum_{j \in p} \chi(p_j, e_i)
\]

where \( b \) encodes the discretization of the PAD position to the corresponding mood-octant (-b corresponds to the inverse octant), \( p_j \) is the \( j \)th parameter of the character’s FFM traits, and \( \chi \) is the mapping between personality and emotion according to [21]. See Table 3 in the Supplemental Material for the full mapping. Note that the emotion generation takes place in step 7 of Algorithm 1. While Equation (7) and (8) can be replaced by other formalisms, we emphasize that they capture the main empirically-based influences of various affective components on emotion: (a) the intensity of an experienced emotion that is close to the current mood is strengthened (and weakened if the experienced emotion is far from the current mood) [26]; and (b) an individual’s personality can up or down regulate emotions. Further, by modeling the influence of mood on emotion, we allow the agent’s mood to influence behavior through the agent’s emotional state. For example, if the agent experiences a negative emotion in an already bad mood, that mood might ‘push the agent over the edge’ (by adjusting the intensity of the negative emotion) and drive it to an emotional reaction. Alternatively, the agent might be ‘only’ driven to regulate the emotion, if its mood manages to inhibit the intensity of the emotion.

5.4 Emotion and Motivation

To enable the agent’s emotional state to influence its motivational state in an empirically grounded manner, we turn to Dialectical Behavior Therapy (DBT) [17]. While there are many valid theories that can connect emotions to motivations, DBT was selected because it clearly outlines how each emotion serves the functions of (a) communicating information to ourselves (which we will refer to here as emotional needs) and (b) motivating us for action [17]. For example, shame communicates to us that we consider something about ourselves to be deficient or inferior. Emotion-Focused Therapy (EFT) [15] and DBT encourage individuals to use the information they acquire from their emotions (such as shame) to detect...
Table 1: A mapping between the OCC emotion types and their corresponding Reiss motivations.

| Emotion  | Corresponding Motivations                  |
|----------|-------------------------------------------|
| Joy      | curiosity, social-contact, eating         |
| Distress | social-contact, family, idealism          |
| Resentment| saving, status                            |
| Pity     | idealism                                  |
| Hope     | idealism, curiosity                       |
| Fear     | tranquility, order                        |
| Satisfaction | saving                            |
| Relief   | tranquility, acceptance                   |
| Disappointment | social-contact, family, idealism |
| Pride    | status, honor                             |
| Admiration| social-contact, honor, idealism           |
| Shame    | acceptance, social-contact                |
| Reproach | vengeance                                 |
| Liking   | family, social-contact, romance           |
| Disliking| vengeance, power                          |
| Gratitude| social-contact, idealism                  |
| Anger    | vengeance, power                          |
| Gratification | acceptance, status, honor     |
| Remorse  | acceptance, idealism, tranquility         |
| Love     | family, social-contact, romance           |
| Hate     | vengeance, power                          |

their emotional needs and to take steps to help fulfill them, in order to ultimately reduce their negative emotions.

Theoretical principles underlying empirically supported psychological treatments (e.g., EFT [15] and DBT [17]) indicate that certain emotions are almost universally related to a small set of emotional needs and based on these principles we draw a correspondence between what is implicitly communicated to the agent by its emotions, and the 16 basic desires which drive the agent’s decision making and behavior. Thus, we allow the agent’s emotional state to influence its behavior. Following this correspondence, an emotion that is experienced with sufficient intensity temporarily influences the agent’s motivation-based decision making. For example, an agent might be feeling shame very intensely; the derived emotional need, according to DBT, is to feel accepted by the environment. Thus, we correlate between the acceptance motivational dimension and shame by assigning an increased importance to acceptance, compared to the emotionally regulated baseline, leading to different action selections by the planner. The planner, to satisfy the need, according to DBT, is to feel accepted by the environment. For example, an agent, A0, which has an oppositional personality, the planner chooses to cheat, powered by the heuristic described in Section 4.2. The configuration of this domain file, including all interactions. Note that these agents also differ in their personality profile. Lastly, each agent is accompanied by a fellow agent, A6; while A0’s disposition does not affect the agents in the scenario, future work will explore more interesting multi-agent settings where there is an interaction between the personality and affective state of different agents. These agents, independently, are placed in the following scenario:

**Scenario** The agent, A0, is very hopeful about passing an exam it is about to take; however, it fails miserably which naturally leads to a strong feeling of disappointment. The agent then proceeds to work on a puzzle and is told that it can cheat and complete it in a fraction of the time it would normally take.

We model this scenario, our model, and the pseudo-code described in Algorithm 1 in the programming language Racket [10]. In each system loop, the symbolically represented state of the world is updated. The system is initialized with a domain file which specifies: (1) agent capabilities; (2) agent personality; (3) the objects in the world and their affordances; (4) target and current motivation vectors (m^c and m^d); (5) state-emotion and state-motivation rules; and (6) initial state. The configuration of this domain file, including the initialization of m^c and m^d, determines the agents’ actions in the scenario. E.g., Table 2 shows that A1 chooses its actions according to its specified domain file, ignoring its affective state. Once the influence of the emotional state on the decision making is enabled, the agent’s behavior will differ from the baseline behavior. Finally, the example is run using an implementation of a heuristic-based planner, powered by the heuristic described in Section 4.2. The domain included roughly 100 actions, 60 state-motivation rules, and 10 objects, and the planner formed plans in less than 100ms.

**Personality → Motivation:** Due to A2’s personality traits, in particular its high agreeableness and conscientiousness, the motivational weight vector w assigns higher importance to honor and idealism. Thus, A2 will not choose to cheat as this will result in lower values in those dimensions. However, since A3 has an opposite personality, the planner chooses to cheat since the motivational gain of succeeding outweighs the negative impact of cheating.

Formally, and similarly to Section 5.1, we modify the weight vector w → w′ to influence the motivationally driven heuristic used by the planner. We define an emotional intensity range, [ILB, ILU), within which we influence w, where ILB and ILU, respectively, are the lower and upper bounds of the range (step 17 of Algorithm 1). Note that if the intensity of an emotion exceeds ILU, an emotional reaction is triggered, as discussed in Section 4.3. As the intensity of the emotion decays, its weighting of the motivational dimensions in w′ weakens, i.e. converges back to w. Finally, if an emotion, ei, has an intensity which is within the defined range (i.e., ILB < Ii(ei) < τr), we define the weight vector as follows:

\[
\mathbf{w}'_i = \begin{cases} 
  \mathbf{w}_i'(1 + (I_i(e_i) - I_{LB})) & \text{if } i \in \zeta_M, e_i \\
  \mathbf{w}_i'(1 - (I_i(e_i) - I_{LB})) & \text{if } i \notin \zeta_M, e_i 
\end{cases},
\]

where \(\zeta_M, e_i\) represents the set of corresponding motivations of the emotion \(e_i\), according to the correspondence between emotions and motivation. Thus, the strength of the emotion’s intensity linearly correlates with the weighting of the motivations. The planner in its forward expansion does not account for the dynamic changes as the emotion’s intensity decays, but rather assumes a fixed metric capturing the current weighting and plans with respect to it.

**: 6 MODEL ASSESSMENT**

In our preliminary evaluation, we set out to illustrate that by slightly varying our model’s configuration, it can generate a diversity of behavior. We additionally demonstrate how the agent’s behavior is influenced by the interaction of the model’s various components.

We introduce 6 agents - A1, A2, A3, A4, A5, and A6 - as shown in Table 2. In each agent’s model, we set a different configuration of the interactions between the various affective components, with agent A1’s model including no interactions and agent A6’s model including all interactions. Note that these agents also differ in their personality profile. Lastly, each agent is accompanied by a fellow agent, A6; while A0’s disposition does not affect the agents in the scenario, future work will explore more interesting multi-agent settings where there is an interaction between the personality and affective state of different agents. These agents, independently, are placed in the following scenario:

**Scenario** The agent, A0, is very hopeful about passing an exam it is about to take; however, it fails miserably which naturally leads to a strong feeling of disappointment. The agent then proceeds to work on a puzzle and is told that it can cheat and complete it in a fraction of the time it would normally take.

We model this scenario, our model, and the pseudo-code described in Algorithm 1 in the programming language Racket [10]. In each system loop, the symbolically represented state of the world is updated. The system is initialized with a domain file which specifies: (1) agent capabilities; (2) agent personality; (3) the objects in the world and their affordances; (4) target and current motivation vectors (m^c and m^d); (5) state-emotion and state-motivation rules; and (6) initial state. The configuration of this domain file, including the initialization of m^c and m^d, determines the agents’ actions in the scenario. E.g., Table 2 shows that A1 chooses its actions according to its specified domain file, ignoring its affective state. Once the influence of the emotional state on the decision making is enabled, the agent’s behavior will differ from the baseline behavior. Finally, the example is run using an implementation of a heuristic-based planner, powered by the heuristic described in Section 4.2. The domain included roughly 100 actions, 60 state-motivation rules, and 10 objects, and the planner formed plans in less than 100ms.

**Personality → Motivation:** Due to A2’s personality traits, in particular its high agreeableness and conscientiousness, the motivational weight vector w assigns higher importance to honor and idealism. Thus, A2 will not choose to cheat as this will result in lower values in those dimensions. However, since A3 has an opposite personality, the planner chooses to cheat since the motivational gain of succeeding outweighs the negative impact of cheating.
Table 2: A comparison between the behavior of different agents, varying in their personality profile and in the degree of influence of their personality \((P \rightarrow X)\) and emotion \((E \rightarrow X)\) on the various affective components. \(R\)'s model includes all relations between the components. \(R\) signifies the emotional Reaction mechanism (see Section 4.3) and \(M\) signifies Motivation.

| Agent | Personality | \(P \rightarrow X\) | \(E \rightarrow X\) | Behavior Timeline |
|-------|-------------|---------------------|---------------------|------------------|
| \(A_1\) | Neutral | 0 | 0 | hopeful | FAIL-EXAM \(\rightarrow\) disappointment | SOLVE-PUZZLE |
| \(A_2\) | \(C^+, A^+\) | \{M\} | 0 | hopeful | FAIL-EXAM \(\rightarrow\) disappointment | SOLVE-PUZZLE \(\rightarrow\) EXTREME-DISAPPOINTMENT |
| \(A_3\) | \(C^-, A^-\) | \{M\} | 0 | hopeful | FAIL-EXAM \(\rightarrow\) disappointment | SOLVE-PUZZLE \(\rightarrow\) EXTREME-DISAPPOINTMENT | LASH-OUT-AT-\(A_0\) |
| \(A_4\) | \(E^-, N^+\) | \{M, E\} | \{M\} | hopeful | FAIL-EXAM \(\rightarrow\) extreme-disappointment | SOLVE-PUZZLE \(\rightarrow\) EXTREME-DISAPPOINTMENT |
| \(A_5\) | \(E^-, N^+\) | \{M, E\} | \{M, R\} | hopeful | FAIL-EXAM \(\rightarrow\) extreme-disappointment | SOLVE-PUZZLE \(\rightarrow\) EXTREME-DISAPPOINTMENT | LASH-OUT-AT-\(A_0\) |
| \(A_6\) | \(E^+, N^-\) | \{M, E\} | \{M, R\} | hopeful | FAIL-EXAM \(\rightarrow\) disappointment | SHARE-DISAPPOINTMENT-WITH-\(A_0\) \(\rightarrow\) ...

**Personality \(\rightarrow\) Emotion:** \(R\)'s personality inhibits its experienced disappointment after failing the exam, preventing an emotional reaction. In contrast, \(A_5\), with a more neurotic personality, will lash out at \(A_0\). While both behaviors are reasonable and depend on the emotional reaction threshold, a personality-driven influence on emotion allows us to modify the personality traits of an agent to drastically change its behavior, rather than remodeling the agent’s domain. E.g., by assigning \(A_5\) a neurotic personality, it experiences negative emotions more intensely, leading to an increased number of emotional reactions on average. Here, the agent’s disappointment led the agent to believe that the exam is unfair and is thus connected to anger and its accompanying emotional reactions.

**Personality \(\rightarrow\) Mood:** The influence of personality on mood, as discussed in Section 5.2, causes the agent’s mood to gravitate towards its personality-determined default mood. Thus, two agents with disparate personalities - e.g., \(A_5\) and \(A_6\) - will have different mood trajectories given their default mood. For example, \(A_6\)’s high extraversion translates to a relatively positive default mood; thus, \(A_6\)’s mood will recover faster after failing the exam compared to \(A_5\), who has a negative default mood due to its high neuroticism.

**Emotion \(\rightarrow\) Motivation:** After failing the exam, \(A_6\)’s disappointment influences the planner, causing it to address the negative emotion, as described in Section 5.4. Since the intensity of the disappointment is between \(I_{LB}\) and \(r_e\), social-contact, family, and idealism are given additional weight in \(w\). The planner, optimizing \(d(m^d, m^e, w)\), plans for \(A_6\) to share its disappointment with \(A_0\).

**Emotion \(\rightarrow\) Emotional Reaction:** As discussed in Section 4.3, we occasionally wish for the agent to be able to immediately express its emotional state, even if that means interrupting what it is currently doing. Since the emotional reaction mechanism \((R\) in Table 2) is disabled in \(A_1\)’s model, it will not lash out at \(A_0\) even when the intensity of its disappointment exceeds \(r_e\) after failing the exam. In some cases, an emotional reaction is expected and is conducive to creating a relatable interaction with the user. E.g., \(A_5\) lashes out at \(A_0\), which is a reaction generated by the emotional reaction planner.

**7 CONCLUSION**

In this work, we introduced a computational model of affect which comprises different affective components - personality, motivation, mood, and emotion - and an empirically-based interplay between these components. Psychological literature guided the formulation of the relations between the components. We model the agent’s affective state and decision making such that the latter is not only influenced by the agent’s motivations, but also by the agent’s affective state. We show, via a series of simulations, that the degree of interaction between affective components in the model, as well as different configurations of the agent’s personality profile, lead to disparate agent behavior. Thus, we justify the need for each of the components and the empirically-based interplay between them. In the future, we will conduct a study to evaluate interactions between our agent and a human. Our model’s ability to address negative emotions could be evaluated using an approach similar to Gratch and Marsella [14], who evaluated an affective model by using a clinical instrument, typically used to assess coping in humans.

Authoring complex and interesting domains for the agent remains a challenge, due to the rule-based nature of our architecture. However, various configurations of our underlying affective model can be reused to create different agents. E.g., one need only change the agent’s personality profile to achieve an entirely different interaction with the agent, reflected in disparate behavior. Since the rules are built on top of well established psychological theories of affect, they can be authored in a relatively general way (e.g., compared to a behaviour tree). Further, while the rules work in a discrete space, their application causes dynamic fluctuations in continuous values (e.g., motivation and emotion) which subsequently lead to dynamic and subtle changes in behavior.

Finally, we emphasize that our proposed model of affect is decoupled from the specific approach for decision making, i.e., optimizing the objective function defined in the motivation space. In our model, the agent’s emotions influence its motivations by increasing or decreasing the weights of the various motivational dimensions, thus dynamically changing the objective function. Optimizing this changing objective then yields the desired behavior. While we chose a heuristic-based forward search planner to greedily optimize the motivations, one can easily replace this with other (possibly more scalable) approaches.

**ACKNOWLEDGEMENTS**

The work presented in this paper was performed at Disney Research Los Angeles in collaboration with DisneyResearchStudios. Shvo thanks Tara Grahnick for helpful discussions surrounding this work.
REFERENCES

[1] Ruth S Aylett, Sandy Louchart, Joao Dias, Ana Paiva, and Marco Vala. 2005. FearNot!—An Experiment in Emergent Narrative. In International Workshop on Intelligent Virtual Agents. Springer, 305–316.

[2] Christian Becker-Asano and Ipke Wachsmuth. 2010. Affective computing with primary and secondary emotions in a virtual human. Autonomous Agents and Multi-Agent Systems 20, 1 (2010), 32.

[3] Tibor Bosse, Matthias Pontier, and Jan Treur. 2007. A dynamical system modeling approach to Gross’ model of emotion regulation. In Proceedings of the 8th international conference on cognitive modeling, ICM, Vol. 7. Taylor and Francis/Psychology Press, Oxford, UK, 187–192.

[4] Sara E Boyd. 2010. A comparison of the Reiss profile with the NEO-PI-R assessment of personality. (2010).

[5] Justine Cassell, Joseph Sullivan, Elizabeth Churchill, and Scott Prevost. 2000. Embodied conversational agents. MIT press.

[6] Edgar Cebolledo and Olga De Troyer. 2015. iATTAC: a system for autonomous agents and dynamic social interactions—the architecture. In Joint International Conference on Serious Games. Springer, 135–146.

[7] Cristina Conati. 2002. Probabilistic assessment of user’s emotions in educational games. Applied artificial intelligence 16, 7–8 (2002), 555–575.

[8] Filipa Ferrada and Luis M Camarinha-Matos. 2017. A system dynamics and agent-based approach to model emotions in collaborative networks. In Doctoral Conference on Computing, Electrical and Industrial Systems. Springer, 29–43.

[9] Richard E Fikes and Nils J Nilsson. 1971. STRIPS: A new approach to the application of theorem proving to problem solving. Artificial intelligence 2, 3–4 (1971), 189–208.

[10] Matthew Flatt. 2010. PLT. Reference: Racket. Technical Report. Technical Report PLT-TR-2010-1, PLT Inc., 2010. http://racket-lang.org/tr1.

[11] Patrick Gebhard. 2005. ALMA: a layered model of affect. In Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems. ACM, 29–36.

[12] Patrick Gebhard, Tanja Schneeberger, Tobias Baur, and Elisabeth André. 2018. MARSSI: Model of Appraisal, Regulation, and Social Signal Interpretation. In Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems. International Foundation for Autonomous Agents and Multiagent Systems, 497–506.

[13] Jonathan Gratch. 2000. Emule: Marshalling passions in training and education. In Proceedings of the fourth international conference on Autonomous agents. ACM, 325–332.

[14] Jonathan Gratch and Stacy Marsella. 2005. Evaluating a computational model of emotion. Autonomous Agents and Multi-Agent Systems 11, 1 (2005), 23–43.

[15] Leslie S Greenberg. 2015. Emotion-focused therapy: Coaching clients to work through their feelings. American Psychological Association.

[16] Mei Yu Lim, João Dias, Ruth Aylett, and Ana Paiva. 2012. Creating adaptive affective autonomous NPCs. Autonomous Agents and Multi-Agent Systems 24, 2 (2012), 287–311.

[17] Marsha Linehan. 2014. DBT Skills Training Manual. Guilford Publications.

[18] Thomas W Malone. 1981. Toward a theory of intrinsically motivating instruction. Cognitive science 5, 4 (1981), 333–369.

[19] Stacy Marsella, Jonathan Gratch, Paolo Petta, et al. 2010. Computational models of emotion. A Blueprint for Affective Computing-A sourcebook and manual 11, 1 (2010), 21–46.

[20] Stacy C Marsella and Jonathan Gratch. 2009. EMA: A process model of appraisal dynamics. Cognitive Systems Research 10, 1 (2009), 70–90.

[21] Robert R McCrae and Paul T Costa. 2003. Personality in adulthood: A five-factor theory perspective. Guilford Press.

[22] Albert Mehrabian. 1978. Measures of individual differences in temperament. Educational and Psychological Measurement 38, 4 (1978), 1105–1117.

[23] Albert Mehrabian. 1996. Analysis of the big-five personality factors in terms of the PAD temperament model. Australian journal of Psychology 48, 2 (1996), 86–92.

[24] Albert Mehrabian. 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology 14, 4 (1996), 261–292.

[25] Albert Mehrabian and Eric O’Reilly. 1980. Analysis of personality measures in terms of basic dimensions of temperament. Journal of Personality and Social Psychology 38, 3 (1980), 492.

[26] Roland Neumann, Beate Seibt, and Fritz Strack. 2001. The influence of mood on the intensity of emotional responses: Disentangling feeling and knowing. Cognition & Emotion 15, 6 (2001), 725–747.

[27] Kenneth R Olson and Dale A Weber. 2004. Relations between big five traits and fundamental motives. Psychological Reports 95, 3 (2004), 795–802.

[28] Andrew Ortony, Gerald L Clore, and Allan Collins. 1990. The cognitive structure of emotions. Cambridge university press.

[29] Rosalind Wright Picard et al. 1995. Affective computing. (1995).

[30] Ricardo Pereira Ramos, Rogério Eduardo da Silva, and Juliane Cristine Koerber Reis. 2012. A Personality Model based on Reiss Motivational Profile for Autonomous Digital Actors. (2012).

[31] Steven Reiss. 2004. Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. Review of general psychology 8, 3 (2004), 179.

[32] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American psychologist 55, 1 (2000), 68.

[33] Maayan Shvo, Jakob Buhmann, and Mubbasir Kapadia. 2019. Modeling the Interaction of Personality, Motivation, Emotion, and Mood in Social Agents (Extended Abstract). In Proceedings of the 8th International Conference on Autonomous Agents and Multi-Agent Systems. International Foundation for Autonomous Agents and Multiagent Systems, 10–17.

[34] Meri Si. 2010. Thespian: a decision-theoretic framework for interactive narratives. University of Southern California.

[35] Ron Sun. 2009. Motivational representations within a computational cognitive architecture. Cognitive Computation 1, 1 (2009), 91–103.