PERSEED: a Self-based Model of Personality for Virtual Agents Inspired by Socio-cognitive Theories

Caroline Faur, Céline Clavel
LIMSI-CNRS
Orsay, France
{faur, clavel}@limsi.fr

Sylvie Pesty
LIG
Grenoble, France
Sylvie.Pesty@imag.fr

Jean-Claude Martin
LIMSI-CNRS
Orsay, France
martin@limsi.fr

Abstract—Consistency and coherence in behaviors are key concepts for creating believable virtual agents. Personality is described by psychology as a set of stable and individualized behaviors. Thus, endowing agents with a personality can help agents to exhibit consistent behaviors. In this paper, we suggest using a socio-cognitive perspective to build a model of personality for virtual agents and to highlight some of the personality structures and processes. Because many researchers in personality psychology emphasize the role of the self in personality, we propose PERSEED, a conceptual model of personality that is based on self-regulation theories.

Keywords—virtual agents; personality; socio-cognitive theories; self-regulation

I. INTRODUCTION

Artificial companions are technological devices that provide services to users in their everyday life on a long-term basis and that elicit the development of a relationship between them and the users [51]. In this regard, artificial companions are included in the user’s social environment. People already respond to computers in a social way [34] but the capacity of companions to behave in a social manner is central to their integration in the social environment. In this paper, we will focus on virtual agents as a subset of artificial companions.

This definition of an artificial companion raises the issue of acceptability. A system is acceptable when users form behavioral intentions of using the system [50]. How could we integrate those technologies in our life as smoothly as possible? Because companions are social actors, people will have expectations about companions’ behaviors. For this reason, artificial companions must be believable. Believability is assessed by the consistency and the coherence of an artificial entity at various levels (psychological and physical; intrapersonal as well as social) [22][36]. We think that social and behavioral believability enhances acceptability. Thus, how can we increase the believability in a way that meets requirements for a companion? How can we help the user in his need to understand a companion’s behaviors?

II. RELATED STUDIES

Personality is a broad concept and psychology sees its study at different levels. In the words of Revelle [41], we can distinguish three levels:
• “All people are the same”: regarding the human species, human exhibit typical behaviors
• “Some people are the same”: this statement refers to the study of individual similarities and differences
• “No person is the same”: each person produces a unique pattern of behavior

Identifying why “some people are the same” gives us insights about elements that should be parts of a model of personality. Investigating why “no person is the same” will help us to create artificial agents that will be perceived more as individuals than as technological objects.

A. Trait-based approach to personality

1) In psychology

Traits are dimensions along which a person can be described. Trait-based theories of personality propose different taxonomies of personality traits. Those theories can be distinguished on two points: 1) the grounding of these dimensions and 2) the number of necessary and sufficient dimensions to describe the personality.

Some researchers based their studies on biological substrates of personality. With this psychobiological background, Eysenck [14], Zuckerman [52] or Cloninger [9] proposed different models, with respectively three, five or seven dimensions. Other psychologists follow the lexical hypothesis. The lexical hypothesis presents two postulates: 1) salient and relevant personality differences will be conveyed by natural language; 2) the most important characteristics are more likely to be expressed by a single word [1][17]. Consequently, the personality vocabulary can be processed with factor analysis to bring out important traits of personality. With Allport’s vocabulary taxonomy, Cattell [7] provided a model with sixteen factors of personality. Using Goldberg’s taxonomy, Costa and McCrae [10] created the Five Factors Model (FFM). The FFM is widely used in studies about individual differences, in psychology and in affective computing. The FFM is drawn on five dimensions of personality, which are also referred to as the Big Five. These five dimensions are: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism.

2) In affective computing

Traits are largely used to model artificial personality, especially the Big Five. Usually, the dimensions of personality are translated into numerical values. Researchers often pick out a subset of traits that are relevant to issues that are addressed by their model. Extraversion, agreeableness and neuroticism form a popular subset because these traits are related to emotions and social behaviors [2][26]. Then, the chosen dimensions are used as weighting parameters in different functionalities. In emotional frameworks for virtual characters, personality has been used to define basic mood [16][26], decay of emotional states [46] and to modulate appraisal process [2]. For conversational agents, personality has been used to shape the style of the discourse [13][49], as well as non-verbal behaviors such as self-adaptors, i.e. non-signaling gestures during speech (e.g. neck scratch or forehead rub) [35]. Personality has also been used to modulate behaviors at a decision-making level by steering goals [2][28][49] or defining preferences in reasoning [46].

The trait-based approach to personality presents some limitations. Virtual characters presented above successfully produce differentiated behaviors but computer scientists have to choose by themselves how to link traits to behaviors. Indeed, traits models are originally designed to provide a description of the personality in terms of broad categories. They focus on the “what” of the personality structure, rather than on “how” or “why” personality leads to specific behaviors [41]. Moreover, trait models hide intra-individual differences. For example, some people behave as extraverted with their friends but show introverted characteristics in their professional environment. Traits do not give us insights into these behaviors because situational features are not accounted for.

B. Socio-cognitive approach to personality

1) In psychology

The socio-cognitive approach to personality underlines the importance of a situation in exhibiting personality behaviors [4]. This approach attempts to understand cognitive and social processes that lead to personality. For that purpose, it focuses on the interaction between the person and the social context (“person X situation” paradigm, also relevant for emotional issues [27]) and highlights the intra-individual differences. Knowing that people show variability in their behaviors across situations, the perceived consistency of personality must be sought in the stability of the behavioral signatures [31]. Understanding these signatures can bring us to understand the mechanisms that are underneath personality behaviors.

A major contributor of this approach is Mischel who, with Shoda, designed the CAPS system (Cognitive Affective Processing System) [31]. CAPS is a meta-theoretical framework of personality. CAPS states that the personality system is characterized by a stable network of cognitive-affective units, which are connected by activation and inhibition links. Behaviors are constructed as a result of the activation spread by situational cues inside this network of mediating units. Five types of cognitive-affective units are defined: encodings, expectancies/beliefs, affects, goals/values and competencies/self-regulatory plans. These categories are broad. The units are purposely left unspecified by the authors. KAPA (Knowledge-and-Appraisal Personality Architecture) [8] is a specification of CAPS. KAPA defines six types of socio-cognitive variables. These variables are divided between structure variables, i.e. knowledge, and process variables, i.e. appraisal. This knowledge/appraisal distinction is crosscut with a direction-of-fit distinction, which distinguished beliefs, goals and evaluative standards.

2) In affective computing

In this perspective, models providing behavior modulators on the basis of situation evaluation can give rise to personality. We focus here on models that are explicitly interested in modeling personality processes that integrate situations. Moffat created Will, an emotion-based personality architecture [32]. By affecting the memory, concerns and
appraisal processes, Moffat stated that Will exhibited some of the Big 5 traits. Sandercock et al. [45] provided a cognitive-affective framework in which personality, represented by coping preferences and weights in appraisal processes, is developed according to the environment. In the BASIC model [43], social cognitive factors are used in combination with Big 5 traits to determine the emotional state of the agents. Neural networks have also been used to model personality in a socio-cognitive perspective: with a situational layer composed of goal features and resource features [40] or with a combination of nodes that represent the Big 5 traits with nodes that represent situational factors [38].

The socio-cognitive approach is seldom used for modeling personality in virtual agents. Most of the presented models are used in perception studies (i.e. a user is looking at an interaction between agents) rather than in an interaction with users. Furthermore, these models are not well adapted to our goal of illustrating underlying processes because they do not focus on specific processes that are described in the psychology literature.

C. Taking a socio-cognitive perspective

We set two goals for our computational model of personality: 1) highlighting personality structures and processes, which means that we must determine which are the structures and the processes to model; 2) increasing believability for virtual agents, to elicit a long-term relationship with users, which means that we must make a model usable for agents evolving in a social environment. Regarding our first goal, socio-cognitive theories provide models that outline personality structures and processes that make “some people the same”. Thus, we use these theories as guidelines in the choice of processes that can be implemented in a computational model. Concerning our second goal, socio-cognitive theories stress the importance of behavior variability due to situations. Following the socio-cognitive perspective will help us to capture intra-individual differences. However, modeling personality is made more complex because we move from five traits to an intricacy of structures and processes in an interplay. To effectively use this approach, we narrow down the possibilities and focus on a specific concept: the self.

III. USING THEORIES ABOUT SELF TO MODEL PERSONALITY IN ARTIFICIAL COMPANIONS

A. Self and personality

In the model of James [23], the self is split between the “I” (self-as-a-perceiver) and the “Me” (self-as-object of perception). The “Me” can be seen as cognitive representations of what characterizes one’s person. The “I” refers to a form of self-awareness, an internal sense of being. However, Skowronski affirms that there is no such thing as “self” but only differentiated sub-systems that, when interacting, contribute to selfness [47]. No real definition of what is the self can be provided today [25].

In spite of the fuzziness of the definition of self, many researchers emphasize the role of the self in personality [6]. [12][44]. Damian and Robins stated that “in fact, the concept of Self is necessary for a complete understanding of personality processes—the processes that generate and regulate thoughts, feelings, and behaviors” [12]. Furthermore, stable aspects of self-knowledge could contribute to inter-situational consistency in behaviors when elements such as perceived self-efficacy could explain inter-situational variability [8].

The concept of self is seldom considered in affective computing. Yet, using self theories is relevant to our goal of making agents such as “no agent is the same” because self is linked to a sense of individuality. Moreover, enabling self-awareness mechanisms in virtual agents might increase their believability [21].

B. Narrowing down to self-regulation: self-discrepancy theory and regulatory-focus theory

The broad notion of self is interesting but not clear enough to be implemented such as in a computational model. In this study, we limit our scope to self-regulation, which refers to mechanisms that are dedicated to aligning expectations or desires with reality [20]. Self-regulatory processes are well integrated in the CAPS framework [3]. Self-regulatory processes produce also behaviors that can be linked to the FFM [15]. To support our conceptual model of personality for artificial companions, we propose to follow the self-discrepancy theory [18] and its evolution, the regulatory-focus theory [19].

The self-discrepancy theory postulates three domains of the self and two standpoints.

Domains of the self are actual self, ideal self and ought self. The actual self refers to the self-concept of a person P. It defines representations of attributes someone (including P) believes P really possesses. The ideal and ought selves refer to the “self-guides” of P. The ideal self contains representations of attributes that someone (including P) wishes P has. The ideal self is related to gains and rewards: P believes he will benefit from being similar to his ideal self. The ought self contains representations of attributes someone (including P) believes P should possess. The ought self is related to losses and sanctions: P believes he will lose by not conforming to his ought self.

Domains of the self are developed according to two standpoints: P’s own personal standpoint and standpoints of P’s significant others (i.e. people who are meaningful in P’s life). Representations from significant others’ standpoints are beliefs that P attributes to those persons. This approach brings to mind the Theory of Mind (ToM), the capacity to attribute mental states to others [5]. Studies about the ToM are interested in how one builds those beliefs and uses them to explain or predict the behaviors of others. Here, we stay focused on how beliefs about oneself drive one’s own behaviors.

The self-discrepancy theory states that people use ideal and ought selves from different standpoints as “self-guides” to compare to their actual self and adapt their behavior to reduce the discrepancy between those different views of the self. Differences in behaviors could be caused by differences in the contents of these selves but also by differences in the
way that one’s uses these contents. These last differences are clarified by the regulatory-focus theory.

The regulatory-focus theory distinguishes between two self-regulation strategies: promotion-focus (concern with the presence or absence of positive outcomes, gains versus non-gains) and prevention-focus (concern with the presence or absence of negative outcomes, losses versus non-losses). Promotion-focus people would be more prone to using their ideal-selves as guides than prevention-focus people, who would prefer using ought-selves [19].

IV. PROPOSED MODEL

Based on personality psychology and the models of self presented above, we propose PERSEED as a self-based model of personality for artificial agents.

Our model does not make assumptions about the cognitive architecture of the agent. This conceptual model modifies the architecture’s functionalities. The personality module can impact the way that information is perceived or how actions are performed, or it can modify a cognitive functionality (e.g. knowledge representation, action selection, emotion mechanism). We will leave the term “cognitive architecture” unspecified in the proposal below, but, where necessary for understanding, we will elicit expected functions of this architecture in the frame of the example.

A. Ideal and ought selves

Following the self-discrepancy theory, the agent has a collection of self-images from different points of view: its own point of view and those of significant others. For example, in the case of artificial companions for children, the child, as the main user, is significant. We can also add other actors who are important in the environment of the child, such as the child’s parents. The actual self is represented by the current state of knowledge in the cognitive architecture.

For each point of view (POV), the agent has an ideal self (what it wants to be) and an ought self (what it should be). These selves are linked to attributes, which are inspired by components defined in Morf’s dynamic self-regulatory processing framework [33]. Attributes could be:

- Self-knowledge: goals and beliefs about the self
- Intrapersonal self-regulatory processes: specific appraisal modes, perception filters, rules of causal attribution or sensitivities to emotions
- Interpersonal self-regulatory strategies: goal setting and planning mechanisms, specific schemas or action preferences

Fig.1 is an illustration of this organization for a gaming pal agent. In this example, we postulate an architecture that provides a capacity for actions, a goal-based selection of actions, an emotional model and a perception of a user’s emotional state. The agent has 3 different points of view: its own, a child’s point of view and the point of view of the child’s parent. For itself, the agent wants to have a low sensibility to negative emotions; then, this sensitivity is linked to its ideal self. The agent thinks that it should elicit positive emotions in its user, so this goal is linked to the agent’s ought self. This last goal is mandatory for the child (ought self) and desirable for the parent (ideal self). For the child, the agent would ideally make many jokes (ideal self) while for the parent, the agent should provide pedagogical advices during games (ought self).

Attributes are then injected into the agent’s cognitive architecture. Injection is the process that selects attributes and transposes these attributes in the cognitive architecture, thereby altering functionalities such as decision-making processes, action selection, self-evaluation or emotional mechanisms. Injection follows rules that are defined in the personality module. These rules are mapped on psychological ways of using self-related contents. Promotion-focus and prevention-focus strategies are examples of rules that can be used. Different rules can be combined to create complex strategies. The mode of injection (i.e. rules used for injection) can be either statically set or selected by the cognitive architecture in regard to contextual specificities. Injection can occur once at the initialization of the agent or it can be replayed bases on feedback given by the cognitive architecture.

For example, a promotion-focus agent will inject attributes that are linked to ideal selves while a prevention-focus agent will inject attributes that are linked to ought selves. In the same way, a selfish agent could inject mostly attributes that are linked to its own selves while a totally altruistic agent can inject only attributes that are linked to significant others. An agent that has the organization presented in Fig. 1 and that uses a promotion-focused strategy will use only attributes that are linked to ideal selves. The decay function for negative emotions (in the emotional model) will be replaced by a new function provided by the personality module. A high-level goal of eliciting positive emotions in user will be added to the goal library. The action “joke” will be weighted to be more likely to be selected when possible.

V. DISCUSSION AND PERSPECTIVES

We identify two complementary approaches to personality: trait theories and socio-cognitive theories. Trait theories define dimensions along which the personality of an
individual can be described. Compared to other models, we do not use a trait structure because we take a more functionalist point of view. As opposed to trait theories, socio-cognitive theories are interested in the structures and processes involved in personality. Taking a socio-cognitive approach to model artificial personality gives us a better insight into the behaviors produced by an agent. 

However, socio-cognitive theories provide models based on the interplay of multiple variables. Thus, these models are difficult to translate in computational terms. For this reason, we bring out one psychological process to model: self-regulation. Self-related structures and processes are often highlighted as central in socio-cognitive theories. Self is a main concept of individuality. Thus, users might attribute a personality to an agent with self mechanisms more easily.

The core of PERSEED is composed of two parts: 1) a network of selves and attributes (i.e. self-related structures and processes) and 2) injection rules (i.e. rules to select the attributes that will be used). Our model works in conjunction with an architecture that provides cognitive functionalities. The attributes injected in this architecture either add new data or functionality to the architecture or replace existing ones. The injection of an attribute such as a goal will be different regarding the architecture. For example, in a BDI architecture [39], PERSEED will add a new desire and, if required, the corresponding plans and beliefs to the database. In the CLARION architecture [48], PERSEED will add a new explicit goal or a new drive in the motivational subsystem. An ad hoc interface must be written for the specific architecture that is intended to be used. If the programming language of the architecture enables it, attributes representing processes can override existing methods at runtime.

PERSEED allows us to explore how differences in the self-regulatory structures and processes are perceived in terms of virtual agents’ personality profiles. Individual differences are modeled in three ways: 1) the contents of the attributes, 2) the links between selves and attributes, and 3) the mode of injection of attributes. Thus, fine-tuned differences can be set up. Our model is extensible because new attributes and new injection rules can be added without modifying the core of the module or pre-existing configurations. In addition, PERSEED is flexible because this model is designed as an external plug-in rather than an embedded part of a specific cognitive architecture.

In future research, we intend to implement this model with a specific cognitive architecture and integrate it into a virtual agent platform. We plan to use the MARC platform [11] which provide highly realistic head 3D models and can display sequences of subtle facial expressions. For the cognitive architecture, we plan to use BIGRE [42], a BDI-like architecture that has already been integrated with MARC. We aim to use our model in a scenario centered on a companion dedicated to help children alone at home after school. First, we will attempt to validate our model with a gaming task. Users will play a game with an agent during repeated sessions. Personality will impact emotions displayed and the strategy used by the agent. Users will be asked to fill a personality questionnaire (selected from existing tests or specifically designed) for the agent. Thus, we can compare results for agents with and without a personality module. We will also test different personality profiles to see whether these differences are perceived. Later, we will add a pedagogical task to explore the links between roles and expression of personality. We will assess acceptability and believability through questionnaires and interviews.

Evolution of the model is considered in a long-term perspective. In a first phase, the links between selves and attributes will be established by developers on the basis of an existing corpus and data in the literature. But, from a developmental viewpoint, personality is shaped by internalized life story [29]. So, links between selves and attributes could evolve by themselves, depending on the repeated interactions between an agent and its user. Moreover, the relationship appearing between a user and a companion is part of the situation. So, social dynamics related to the relationship, such as dominance or warmth, will be integrated as an entry of the model. Furthermore, virtual agents are not the only type of artificial companions. We intend to extend this model to robots to provide a unified framework to model personality for artificial companions.

ACKNOWLEDGMENT
This work was supported by the French National Agency for Research (project reference: ANR-12-CORD-019).

REFERENCES