



# Adaptation Mechanisms in Human-Agent Interaction: Effects on User's Impressions and Engagement

Beatrice Biancardi<sup>1\*</sup>, Soumia Dermouche<sup>2</sup> and Catherine Pelachaud<sup>2</sup>

<sup>1</sup>LTCl, Télécom Paris, Institut Polytechnique de Paris, Paris, France, <sup>2</sup>CNRS-ISIR, Sorbonne University, Paris, France

## OPEN ACCESS

### Edited by:

Stefano Triberti,  
University of Milan, Italy

### Reviewed by:

Sandra Cano,  
Pontificia Universidad Católica de  
Valparaíso, Chile  
Benjamin Lok,  
University of Florida, United States

### \*Correspondence:

Beatrice Biancardi  
beatrice.biancardi@telecom-paris.fr

### Specialty section:

This article was submitted to  
Human-Media Interaction,  
a section of the journal  
Frontiers in Computer Science

**Received:** 17 April 2021

**Accepted:** 07 July 2021

**Published:** 12 August 2021

### Citation:

Biancardi B, Dermouche S and  
Pelachaud C (2021) Adaptation  
Mechanisms in Human-Agent  
Interaction: Effects on User's  
Impressions and Engagement.  
Front. Comput. Sci. 3:696682.  
doi: 10.3389/fcomp.2021.696682

Adaptation is a key mechanism in human-human interaction. In our work, we aim at endowing embodied conversational agents with the ability to adapt their behavior when interacting with a human interlocutor. With the goal to better understand what the main challenges concerning adaptive agents are, we investigated the effects on the user's experience of three adaptation models for a virtual agent. The adaptation mechanisms performed by the agent take into account the user's reaction and learn how to adapt on the fly during the interaction. The agent's adaptation is realized at several levels (i.e., at the behavioral, conversational, and signal levels) and focuses on improving the user's experience along different dimensions (i.e., the user's impressions and engagement). In our first two studies, we aim to learn the agent's multimodal behaviors and conversational strategies to dynamically optimize the user's engagement and impressions of the agent, by taking them as input during the learning process. In our third study, our model takes both the user's and the agent's past behavior as input and predicts the agent's next behavior. Our adaptation models have been evaluated through experimental studies sharing the same interacting scenario, with the agent playing the role of a virtual museum guide. These studies showed the impact of the adaptation mechanisms on the user's experience of the interaction and their perception of the agent. Interacting with an adaptive agent vs. a nonadaptive agent tended to be more positively perceived. Finally, the effects of people's *a priori* about virtual agents found in our studies highlight the importance of taking into account the user's expectancies in human-agent interaction.

**Keywords:** human-agent interaction, adaptation mechanisms, engagement, impressions, embodied conversational agent (ECA)

## 1 INTRODUCTION

During an interaction, we communicate through multiple behaviors. Not only speech but also our facial expressions, gestures, gaze direction, body orientation, etc. participate in the message being communicated (Argyle, 1972). Both interactants are active participants in an interaction and adapt their behaviors to each other. This adaptation arises on several levels: we align ourselves linguistically (vocabulary, syntax, and level of formality), but we also adapt our nonverbal behaviors (e.g., we respond to the smile of our interlocutor, and we imitate their posture and their gestural expressiveness), our conversational strategies (e.g., to be perceived as warmer or more competent), etc. (Burgoon et al., 2007). This multilevel adaptation can have several functions, such as reinforcing engagement in the interaction, emphasizing our relationship with others, showing empathy, and managing the impressions we give to others (Lakin and Chartrand, 2003;

Gueguen et al., 2009; Fischer-Lokou et al., 2011). The choice of verbal and nonverbal behaviors and their temporal realization are markers of adaptation.

Embodied conversational agents (ECAs) are virtual entities with a humanlike appearance that are endowed with communicative and emotional capabilities (Cassell et al., 2000). They can display a wide range of multimodal expressions to be active participants in the interaction with their human interlocutors. They have been deployed in various human-machine interactions where they can act as a tutor (Mills et al., 2019), health support (Lisetti et al., 2013; Rizzo et al., 2016; Zhang et al., 2017), a companion (Sidner et al., 2018), a museum guide (Kopp et al., 2005; Swartout et al., 2010), etc. Studies have reported that ECAs are able to take into account their human interlocutors and show empathy (Paiva et al., 2017), display backchannels (Bevacqua et al., 2008), and build rapport (Huang et al., 2011; Zhao et al., 2016). Given its relevance in human-human interaction, adaptation could be exploited to improve natural interactions with ECAs. It thus seems important to investigate whether an agent adapting to the user's behaviors could provoke similar positive outcomes in the interaction.

The majority of works in this context developed models learnt from existing databases of human-human interaction and did not consider the dynamics of adaptation mechanisms during an interaction. We are interested in developing an ECA that exploits how the interaction is currently going and is able to learn in real time what the best adaption mechanism for the interaction is.

In this article, we report three studies where an ECA adapts its behaviors by taking into account the user's reaction and by learning how to adapt on the fly during the interaction.

The goal of the different studies is to answer two broad research questions:

*“Does adapting an ECA's behavior enhance user's experience during interaction?”*

*“How does an ECA which adapts its behavior in real-time influence the user's perception of the agent?”*

A user's experience can involve many factors and can be measured by different dimensions, such as the user's engagement and the user's impressions about the ECA (Burgoon et al., 2007). In our three studies that we report in this article, we implemented three independent models where the agent's adaptation is realized at several levels and focuses on improving the user's experience along different dimensions as follows:

- 1) the agent's adaptation at a behavioral level: the ECA adapts its behaviors (e.g., gestures, arm rest poses, and smiles) in order to maximize the user's impressions about the agent's warmth or competence, the two fundamental dimensions of social cognition (Fiske et al., 2007). This model is described in **Section 7**;
- 2) the agent's adaptation at a conversational level: the ECA adapts its communicative strategies to elicit different levels

of warmth and competence, in order to maximize the user's engagement. This model is described in **Section 8**; and

- 3) the agent's adaptation at a signal level: the ECA adapts its head and eye rotation and lip corner movement in function of the user's signals in order to maximize the user's engagement. This model is described in **Section 9**.

Each adaptation mechanism has been implemented in the same architecture that allows an ECA to adapt to the nonverbal behaviors of the user during the interaction. This architecture includes a multimodal analysis of the user's behavior using the Eyesweb platform (Camurri et al., 2004), a dialogue manager (Flipper (van Waterschoot et al., 2018)), and the ECA GRETA (Pecune et al., 2014). The architecture has been adapted to each model and evaluated through experimental studies. The ECA played the role of a virtual guide at the Science Museum of Paris. The scenario used in all the evaluation studies is described in **Section 6**.

Even though these three models have been implemented in the same architecture and tested on the same scenario, they have not been developed in order to do comparative studies. The main goal of this paper is to frame them in the same theoretical framework (see **Section 2**) and have insights into each of these different adaptation mechanisms to better understand what the main challenges concerning these models are and to suggest further improvements for an adaptation system working on multiple levels.

This article is organized as follows: in **Section 2**, we review the main theories about adaptation which our work relies on, in particular Burgoon and others' work; in **Section 3**, we present an overview of existing models that focus on adapting the ECA's behavior according to the user's behavior; in **Section 4**, we specify the dimensions we focused on in our adaptation models; in **Section 5**, we present the general architecture we conceived to endow our ECA with the capability of adapting its behavior to the user's reactions in real time; in **Section 6**, we describe the scenario we conceived to test the different adaptation models; in **Sections 7–9**, we report the implementation and evaluation of each of the three models. More details about them can be found in our previous articles (Biancardi et al., 2019b; Biancardi et al., 2019a; Dermouche and Pelachaud, 2019). We finally discuss the results of our work and possible improvements in **Sections 10, 11**, respectively.

## 2 BACKGROUND

Adaptation is an essential feature of interpersonal relationships (Cappella, 1991). During an effective communication, people adapt their interaction patterns to one another's (e.g., dancers synchronize their movements and people adapt their conversational style in a conversation). These patterns contribute to defining and maintaining our interpersonal relationships, by facilitating smooth communication, fostering attraction, reinforcing identification with an in-group, and increasing rapport between communicators (Bernieri et al.,

1988; Giles et al., 1991; Chartrand and Bargh, 1999; Gallois et al., 2005).

There exist several adaptation patterns, differing according to their behavior type (e.g., the modality, the similarity to the other interlocutor's behavior, etc.), their level of consciousness, whether they are well decoded by the other interlocutor, and their effect on the interaction (Toma, 2014). Cappella and others (Cappella, 1981) considered an additional characteristic, that is, adaptation can be asymmetrical (unilateral), when only one partner adapts to the other, or symmetrical (mutual), like in the case of interaction synchrony.

In line with these criteria, in some examples of adaptation, people's behaviors become more similar to one another's. This type of adaptation is often unconscious and reflects reciprocity or convergence. According to Gouldner (Gouldner, 1960), reciprocity is motivated by the need to maintain harmonious and stable relations. It is contingent (i.e., one person's behaviors are dependent upon the other's) and transactional (i.e., it is part of an exchange process between two people).

In other cases, adaptation can include complementarity or divergence; this occurs when the behavior of one person differs from but complements that of the other person.

Several theories focus on one or more specific characteristics of adaptation and highlight different factors that drive people's behaviors. They can be divided into four main classes according to the perspective they follow to explain adaptation.

The first class of theories includes biologically based models (e.g., (Condon and Ogston, 1971), (Bernieri et al., 1988)). These theories state that individuals exhibit similar patterns to one another. These adaptation patterns have an innate basis, as they are related to satisfaction of basic needs like bonding, safety, and social organization. Their innate bases make them universal and involuntary, but they can be influenced by environmental and social factors as well.

Following a different perspective, arousal-based and affect-based models (e.g., (Argyle and Dean, 1965), (Altman et al., 1981), (Cappella and Greene, 1982)) support the role of internal emotional and arousal states as driving factors of people's behaviors. These states determine approaching or avoiding behaviors. This group of theories explains the balance between compensation and reciprocity.

Social-norm models (e.g., (Gouldner, 1960), (Dindia, 1988)) do not consider the role of physiological or psychological factors but argue for the importance of social phenomena as guiding forces. These social phenomena are, for example, the in-group or out-group status of the interactants, their motivation to identify with one another, and their level of affiliation or social distance.

The last class of theories includes communication- and cognition-based models (e.g., (Andersen, 1985), (Hale and Burgoon, 1984)), which focus on the communicative purposes of the interactants and on the meaning that the behavioral patterns convey. While adaption happens mainly unconsciously, it may happen that the process of interpersonal adaptation may be strategic and conscious (Giles et al., 1991; Gallois et al., 2005).

The majority of these theories have been studied by Burgoon and others (Burgoon et al., 2007). In particular, they examined

fifteen previous models and considered the most important conclusions from the previous empirical research. From this analysis, they came out with a broader theory, the interaction adaptation theory (IAT). This theory states that we alter our behavior in response to the behavior of another person in conversations (Infante et al., 2010). IAT takes into account the complexities of interpersonal interactions by considering people's needs, expectations, desires, and goals as precursors of their degree and form of adaptation. IAT is a communication theory made of multiple theories, which focuses on the sender's and the receiver's process and patterns.

Three main interrelated factors contribute to IAT. Requirements (Rs) refer to the individual beliefs about what is necessary in order to have a successful interaction. Rs are mainly driven by biological factors, such as survival, safety, and affiliation. Expectations (Es) refer to what people expect from the others based on social norms or knowledge coming from previous interactions. Es are mainly influenced by social factors. Finally, desires (Ds) refer to the individual's goals and preferences about what to get out of the interaction. Ds are mainly influenced by person-specific factors, such as temperament or cultural norms. These three factors are used to predict an individual's interactional position (IP). This variable derives from the combination of Rs, Es, and Ds and represents the individual's behavioral predisposition that will influence how an interaction will work. The IP would not necessarily correspond to the partner's actual behavior performed in the interaction (A). The relation between the IP and A will determine the type of adaptation during the interaction. For example, when the IP and A almost match, IAT predicts behavioral patterns such as reciprocity and convergence. When A is more negatively valenced than the IP, the model predicts compensation and avoiding behaviors.

In the work presented in this article, we rely on Burgoon's IAT. Indeed, our adapting ECA has an interactional position (IP), resulting from its desires (Ds) and expectations (Es). In particular, the agent's desire (D) is to maximize the user's experience, and its expectations (Es) are about the user's reactions to its behaviors. In our different models of adaptation mechanisms, the agent's desire (D) refers either to giving the best impression to the user or to maximizing the user's engagement (see **Section 4**). Consequently, the expectations (Es) refer to the user's reaction reflecting their impressions or engagement level in response to the agent's behavior. The behavior that will be performed by the ECA depends on the relation between the agent's IP and the user's reaction (actual behavior A).

In addition, we explore different ways in which the ECA can adapt to the user's reactions. On one hand, we focus on theories that consider adaptive behaviors more broadly than a mere matching, that is, adaptation as responding in appropriate ways to a partner. The ECA will choose its behaviors according to the effect they have on the user's experience (see **Section 7**). In Study 2 (see **Section 8**), our adaptive agent follows the same perspective but by adapting its communicative strategies. On the other hand, we try to simulate a more unconscious and automatic process working at a motoric level; the agent adapts at a signal level (see Study 3, **Section 9**).

### 3 STATE OF THE ART

In this section, we present an overview of existing models that focused on adapting ECAs' behavior according to the user's behavior in order to enhance the interaction and the user's experience along different dimensions such as engagement, rapport, interest, liking, etc. These existing models predicted and generated different forms of adaptation, such as backchannels, mimicry, and voice adaptation, and were applied on virtual agents or robots.

Several works were interested in understanding the impact of adaptation on the user's engagement and rapport building. Some of them did so through the production of backchannels. Huang et al. (2010) developed an ECA that was able to produce backchannels to reinforce the building of rapport with its human interlocutor. The authors used conditional random fields (CRFs) (Lafferty et al., 2001) to automatically learn when listeners produce visual backchannels. The prediction was based on three features: prosody (e.g., pause and pitch), lexical (spoken words), and gaze. Using this model, the ECA was perceived as more natural; it also created more rapport with its interlocutor during the interaction. Schröder et al. (2015) developed a sensitive artificial listener that was able to produce backchannels. They developed a model that predicted when an ECA should display a backchannel and with which intention. The backchannel could be either a smile, nod, and vocalization or an imitation of a human's smile and head movement. Participants who interacted with an ECA displaying backchannels were more engaged than they were when no backchannels were shown.

Other works focused on modeling ECAs that were able to mimic their interlocutors' behaviors. Bailenson and Yee (2005) studied the social influence of mimicry during human-agent interaction (they referred to this as the chameleon effect). The ECA mimicked the user's head movements with a delay of up to 4 s. An ECA showing mimicry was perceived as more persuasive and more positive than an ECA showing no mimicry at all. Raffard et al. (2018) also studied the influence of ECAs mimicking their interlocutors' head and body posture with some delay (below 4 s). Participants with schizophrenia and healthy participants interacted with an ECA that either mimicked them or not. Both groups showed higher behavior synchronization and reported an increase in rapport in the mimicry condition. Another study involving mimicry was proposed by (Verberne et al., 2013) in order to evaluate if an ECA mimicking the user's head movements would be liked and trusted more than a non-mimicking one. This research question was investigated by running two experiments in which participants played a game involving drivers handing over the car control to the ECA. While results differed depending on the game, the authors found that liking and trust were higher for a mimicking ECA than for a non-mimicking one.

Reinforcement learning methods for optimizing the agent's behaviors according to the user's preference have been used in different works. For example, Liu et al. (2008) endowed a robot with the capacity to detect, in real time, the affective states (liking, anxiety, and engagement) of children with autism spectrum disorder and to adapt its behavior to the children's preferences

of activities. The detection of children's affective states was done by exploiting their physiological signals. A large database of physiological signals was explored to find their interrelation with the affective states of the children. Then, an SVM-based recognizer was trained to match the children's affective state to a set of physiological features. Finally, the robot learned the activities that the children preferred to do at a moment based on the predicted liking level of the children using QV-learning (Wiering, 2005). The proposed model led to an increase in the reported liking level of the children toward the robot. Ritschel et al. (2017) studied the influence of the agent's personality on the user's engagement. They proposed a reinforcement learning model based on social signals for adapting the personality of a social robot to the user's engagement level. The user's engagement was estimated from their multimodal social signals such as gaze direction and posture. The robot adapted its linguistic style by generating utterances with different degrees of extroversion using a natural language generation approach. The robot that adapted its personality through its linguistic style increased the user's engagement, but the degree of the user's preference toward the robot depended on the ongoing task. Later on, the authors applied a similar approach to build a robot that adapts to the sense of humor of its human interlocutor (Weber et al., 2018).

Several works have been conducted in the domain of education where an agent, being physical as a robot or virtual as an ECA, adapted to the learner's behavior. These works reported that adaptation is generally linked with an increase in the learner's engagement and performance. For example, Gordon et al. (2016) developed a robot acting as a tutor for children learning a second language. To favor learning, the robot adapted its behaviors to optimize the level of the children's engagement, which was computed from their facial expressions. A reinforcement learning algorithm was applied to compute the robot's verbal and nonverbal behavior. Children showed higher engagement and learned more second-language words with the robot that adapted its behaviors to the children's facial expression than they did with the nonadaptive robot. Woolf et al. (2009) manually designed rules to adapt the facial expressions of a virtual tutor according to the student's affective state (e.g., frustrated, bored, or confused). For example, if the student was delighted and sad, respectively, the tutor might look pleased and sad, respectively. Results showed that when the virtual tutor adapted its facial expressions in response to the student's ones, the latter maintained higher levels of interest and reduced levels of boredom when interacting with the tutor.

Other works looked at adapting the activities undertaken by an agent during an interaction to enhance knowledge acquisition and reinforce engagement. In the study by (Ahmad et al., 2017), a robot playing games with children was able to perform three different types of adaptations, game-based, emotion-based, and memory-based, which relied, respectively, on the following: 1) the game state, 2) emotion detection from the child's facial expressions, and 3) face recognition mechanisms and remembering the child's performance. In the first category of adaptation, a decision-making mechanism was used to generate a supporting verbal and nonverbal behavior. For example, if the



child performed well, the robot said “Wow, you are playing extraordinary” and showed positive gestures such as a thumbs-up. The emotion-based adaptation mapped the child’s emotions to a set of supportive dialogues. For example, when detecting the emotion of joy, the robot said, “You are looking happy, I think you are enjoying the game.” For memory adaptation, the robot adapted its behavior after recognizing the child and retrieving the child’s game history such as their game performance and results. Results highlighted that emotion-based adaptation resulted in the highest level of social engagement compared to memory-based adaptation. Game adaptation did not result in maintaining long-term social engagement. Coninx et al. (2016) proposed an adaptive robot that was able to change activities during an interaction with children suffering from diabetes. The aim of the robot was to reinforce the children’s knowledge with regard to managing their disease and well-being. Three activities were designed to approach the diabetes-learning problem from different perspectives. Depending on the children’s motivation, the robot switched between the three proposed activities. Adapting activities in the course of the interaction led to a high level of children’s engagement toward the robot. Moreover, this approach seemed promising for setting up a long-term child–robot relationship.

In a task-oriented interaction, Hemminahaus and Kopp (2017) presented a model to adapt the social behavior of an assistive robot. The robot could predict when and how to guide the attention of the user, depending on the interaction contexts. The authors developed a model that mapped interactional functions such as motivating the user and guiding them onto low-level behaviors executable by the robot. The high-level functions were selected based on the interaction context and the attentive and emotional states of the user. Reinforcement learning was used to predict the mapping of these functions onto lower-level behaviors. The model was evaluated in a scenario in which a robot assisted the user in solving a memory game by guiding their attention to the target objects. Results showed that users were able to solve the game faster with the adaptive robot.

Other works focused on voice adaptation during social interaction. Voice adaptation is based on acoustic-prosodic entrainment that occurs when two interactants adapt their manner of speaking, such as their speaking rate, tone, or pitch, to each other’s. Levitan (2013) found that voice adaptation improved spoken dialogue systems’ performance and the user’s satisfaction. Lubold et al. (2016) studied the effect of voice adaptation on social variables such as rapport and social presence. They found that social presence was significantly higher with a social voice-adaptive speech interface than with purely social dialogue.

In most previous works, the adaptation mechanisms that have been implemented measured their influence on the user’s engagement through questionnaires. They did not include them as a factor of the adaptation mechanisms. In our first two studies reported in this article, we aimed to learn the agent’s multimodal behaviors and conversational strategies to dynamically optimize the user’s engagement and their impressions of the ECA, by taking them as input during the learning process.

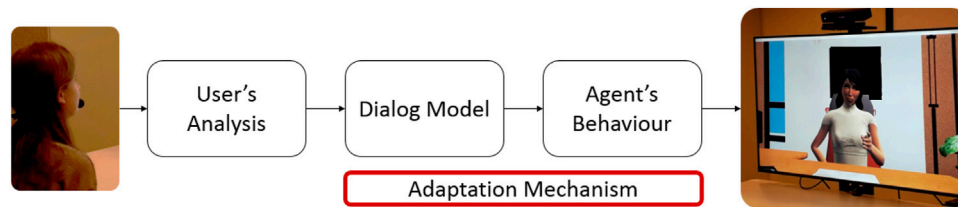
Moreover, in most existing works, the agent’s predicted behavior depended exclusively on the user’s behavior and ignored the interaction loop between the ECA and the user. In our third study, we took into account this interaction loop, that is, our model takes as input both the user’s and the agent’s past behavior and predicts the agent’s next behavior. Another novelty presented in our work is to include the agent’s communicative intentions along with its adaptive behaviors.

## 4 DIMENSIONS OF STUDY

In our studies, we focused on adaptation in human–agent interaction by using the user’s reactions as the input for the agent’s adaptation. In particular, we took into account two main dimensions, which are the user’s impressions of the ECA and the user’s engagement during the interaction.

These two dimensions play an important role during human–agent interactions, as they influence the acceptability of the ECA by the user and the willingness to interact with it again (Bergmann et al., 2012; Bickmore et al., 2013; Cafaro et al., 2016). In order to engage the user, it is important that the ECA displays appropriate socio-emotional behaviors (Pelachaud, 2009). In our case, we were interested in whether and how the ECA could affect the user’s engagement by managing the impressions it gave to them. In particular, we considered the user’s impressions of the two main dimensions of social cognition, that is, warmth and competence (Fiske et al., 2007). Warmth includes traits like friendliness, trustworthiness, and sociability, while competence includes traits like intelligence, agency, and efficacy. In human–human interaction, several studies have showed the role of nonverbal behaviors in conveying different impressions of warmth and competence. In particular, communicative gestures, arm rest poses, and smiling behavior have been found to be associated with different degrees of warmth and/or competence (Duchenne, 1990; Cuddy et al., 2008; Maricchiolo et al., 2009; Biancardi et al., 2017a). In the context of human–agent interaction, we can control and adapt the nonverbal behaviors of the ECA during the flow of the interaction.

Following Burgoon’s IAT theoretical model, our adapting ECA thus has the desire *D* to maintain the user’s engagement (or impressions) during the interaction. Since the ECA aims to be perceived as a social entity by its human interlocutor, the agent’s expectancy *E* is that adaptation can enhance the interaction experience. In our work, we are interested in whether adapting at a behavioral or conversational level (i.e., the agent’s warmth and competence impressions) and/or at a low level (i.e., the agent’s head and eye rotation and lip corner movement) could affect the user’s engagement. Even though the impact of the agent’s adaptation on the user’s engagement has already been the object of much research (see Section 3), here we use the user’s engagement as a real-time variable given as input for the agent’s adaptation.



**FIGURE 1 |** System architecture: in the User's Analysis module, the user's nonverbal and verbal signals are extracted and interpreted and the user's reaction is sent to the Dialogue Model module, which computes the dialogue act to be communicated by the ECA. The Agent's Behavior module instantiates the dialogue act into multimodal behaviors to be displayed by the ECA. The Adaptation Mechanism module adapts the agent's behavior to the user's behavior. Its placement in the architecture depends on the specific adaptation mechanism that is implemented.



**FIGURE 2 |** Interaction space in the experiment room. The participants were sitting in front of the TV screen displaying the ECA. On the left, two screens separated the interaction space from the control space.

## 5 ARCHITECTURE

In this section, we present the architecture we conceived to endow the ECA with the capability of adapting its behavior to the user's reactions in real time. The architecture consists of several modules (see **Figure 1**). One module extracts information about the user's behaviors using a Kinect and a microphone. This information is interpreted in terms of speech (what the user has uttered) and the user's state (e.g., their engagement in the interaction). This interpreted information is sent to a dialogue manager that computes the communicative intentions of the ECA, that is, what it should say and how. Finally, the animation of the ECA is computed on the fly and played in real time. The agent's adaptation mechanisms are also taken into account when computing its verbal and nonverbal behaviors. The architecture is general enough to allow for the customization of its different modules according to the different adaptation mechanisms and goals of the agent.

In more detail, the four main parts of the architecture are as follows:

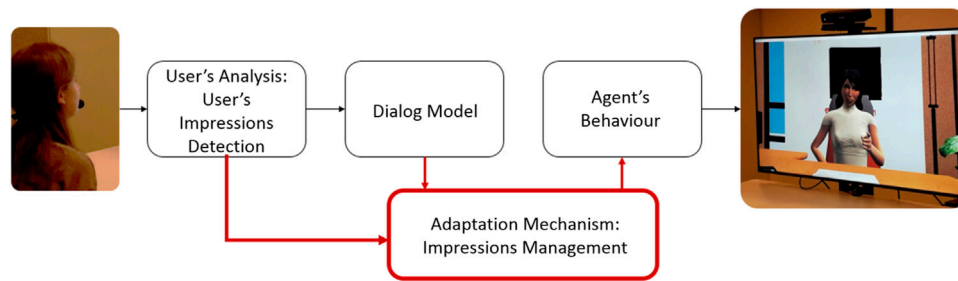
- 1) **User's Analysis:** the EyesWeb platform (Camurri et al., 2004) allows the extraction in real time of the following: 1) the user's nonverbal signals (e.g., head and trunk rotation), starting from

the Kinect depth camera skeleton data; 2) the user's facial muscular activity (action units or AUs (Ekman et al., 2002)), by running the OpenFace framework (Baltrušaitis et al., 2016); 3) the user's gaze; and 4) the user's speech, by executing Microsoft Speech Platform<sup>1</sup>.

These low-level signals are processed using EyesWeb and other external tools, such as machine learning pretrained models (Dermouche and Pelachaud, 2019; Wang et al., 2019), to extract high-level features about the user, such as their level of engagement.

- 2) **Dialogue Model:** in this module, the dialogue manager Flipper (van Waterschoot et al., 2018) selects the dialogue act that the agent will perform and the communicative intention of the agent (i.e., how to perform that dialogue act).
- 3) **Agent's Behavior:** the agent's behavior generation is performed using GRETA, a software platform supporting the creation of socio-emotional embodied conversational agents (Pecune et al., 2014). The Agent's Behavior module is made of two main modules: the Behavior Planner receives the communicative intentions of the ECA from the Dialogue Model module as input and instantiates them into multimodal behaviors and the Behavior Realizer transforms the multimodal behaviors into facial and body animations to be displayed on a graphics screen.
- 4) **Adaptation Mechanism:** since the ECA can adapt its behaviors at different levels, the Adaptation Mechanism module is implemented in different parts of the architecture, according to the type of adaptation that the ECA performs. That is, the adaptation can affect the communicative intentions of the ECA, or it can occur during the behavior realization at the animation level. In the first two models presented in this article, the Adaptation Mechanism module is connected to the Dialogue Model module, while for the third model, it is connected to the Agent's Behavior module.

<sup>1</sup><https://www.microsoft.com/en-us/download/details.aspx?id=27225>



**FIGURE 3 |** Modified system architecture used in Study 1. In particular, the User's Analysis module contains the model to detect the user's impressions from facial signals. The Impressions Management module contains the Q-learning algorithm.

## 6 SCENARIO

Each type of adaptation has been investigated by running human-agent interaction experiments at the Science Museum of Paris. In the scenario conceived for these experiments, the ECA, called Alice, played the role of a virtual guide of the museum.

The experiment room included a questionnaire space, including a desk with a laptop and a chair; an interaction space, with a big TV screen displaying the ECA, a Kinect Two placed on the top of the TV screen, and a black tent behind the chair where the participant sat; and a control space, separated from the rest of the room by two screens, including a desk with the computer running the system architecture. The interaction space is shown in **Figure 2**.

The experiments were completed in three phases as follows:

- 1) before the interaction began, the participant sat at the questionnaire space, read and signed the consent form, and filled out the first questionnaire (NARS, see below). Then they moved to the interaction space, where the experimenter gave the last instructions [5 min];
- 2) during the interaction phase, the participant stayed right in front of the TV screen, between it and the black tent. They wore a headset and were free to interact with the ECA as they wanted. During this phase, the experimenter stayed in the control space, behind the screens [3 min]; and
- 3) after the interaction, the participant came back to the questionnaire space and filled out the last questionnaires about their perception of the ECA and of the interaction. After that, the experimenter proceeded with the debriefing [5 min].

Before the interaction with the ECA, we asked participants to fill out a questionnaire about their *a priori* about virtual characters (NARS); an adapted version of the NARS scale from the study by Nomura et al. (2006) was used. Items of the questionnaire included, for example, how much participants would feel relaxed talking with a virtual agent, or how much they would like the idea that virtual agents made judgments.

The interaction with the ECA lasted about 3 min. It included 26 steps. A step included one dialogue act played by the ECA and the participant's potential reaction/answer. The dialogue scenario

was built so that the ECA drove the discussion. The virtual guide provided information on an exhibit that was currently happening in the museum. It also asked some questions about participants' preferences. Purposely, we limited the possibility for participants to take the lead in the conversation as we wanted to avoid any error due to automatic speech understanding. More details about the dialogue model can be found in the study by (Biancardi et al., 2019a).

## 7 STUDY 1: ADAPTATION OF AGENT'S BEHAVIORS

At this step, we aim to investigate adaptation at a high level, meant as convergence of the agent's behaviors according to the user's impressions of the ECA.

The goal of this first model is to make the ECA learn the verbal and nonverbal behaviors to be perceived as warm or competent by measuring and using the user's impressions as a reward.

### 7.1 Architecture

The general architecture described in **Section 5** has been modified in order to contain a module for the detection of the user's impressions and a specific set of verbal and nonverbal behaviors from which the ECA could choose.

The modified architecture of the system is depicted in **Figure 3**. In the following section, we give more details about the modified modules.

#### 7.1.1 User's Analysis: User's Impression Detection

The user's impressions can be detected from their nonverbal behaviors, in particular, their facial expressions. The User's Analysis module is integrated with a User's Impression Detection module that takes as input a stream of the user's facial action units (AUs) (Ekman et al., 2002) and outputs the potential user's impressions about the level of warmth (or competence) of the ECA.

A trained multilayer perceptron (MLP) regression model is implemented in this module to detect the impressions formed by users about the ECA. The MLP model was previously trained using a corpus including face video recordings and continuous self-report annotations of warmth and competence given by

participants watching the videos of the NoXi database (Cafaro et al., 2017). The self-report annotations being considered separately, the MLP model was trained twice, one for warmth and one for competence. More details about this model can be found in the study by (Wang et al., 2019).

### 7.1.2 Adaptation Mechanism: Impression Management

In this model, the adaptation of the ECA concerns the impressions of warmth and competence given to the user. The inputs of the Adaptation Mechanism module are the dialogue act to be realized (coming from the Dialogue Model module) and the user's impression of the agent's warmth or competence (coming from the User's Analysis module). The output is a combination of behaviors to realize the dialogue act, chosen from a set of possible verbal and nonverbal behaviors to perform.

To be able to change the agent's behavior according to the detected participant's impressions, a machine learning algorithm is applied. We follow a reinforcement learning approach to learn which actions the ECA should take (here, verbal and nonverbal behaviors) in response to some events (here, the user's detected impressions). We rely on a Q-learning algorithm for this step. More details about it can be found in the study by (Biancardi et al., 2019b).

The set of verbal and nonverbal behaviors, from which the Q-learning algorithm selects a combination to send to the Behavior Planner of the Agent's Behavior module, includes the following:

- Type of gestures: the ECA could perform ideational (i.e., related to the content of the speech) or beat (i.e., marking speech rhythm, not related to the content of the speech) gestures or no gestures.
- Arm rest poses: in the absence of any kind of gesture, these rest poses could be performed by the ECA: akimbo (i.e., hands on the hips), arms crossed on the chest, arms along its body, or hands crossed on the table.
- Smiling: during the animation, the ECA could decide whether or not to perform smiling behavior, characterized by the activation of AU6 (cheek raiser) and AU12 (lip puller-up).
- Verbal behavior: the ECA could modify the use of you- and we-words, the level of formality of the language, and the length of the sentences. These features have been found to be related to different impressions of warmth and competence (Pennebaker, 2011; Callejas et al., 2014).

## 7.2 Experimental Design

The adaptation model described in **Subsection 7.1.2** has been evaluated by using the scenario described in **Section 6**. Here, we describe the experimental variables manipulated and measured during the experiment.

### 7.2.1 Independent Variable

The independent variable manipulated in this experiment, called Model, concerns the use of the adaptation model and includes three conditions:

- Warmth: when the ECA adapts its behaviors according to the user's impressions of the agent's warmth, with the goal to maximize these impressions;
- Competence: when the ECA adapts its behaviors according to the user's impressions of the agent's competence, with the goal to maximize these impressions; and
- Random: when the adaptation model is not exploited and the ECA randomly chooses its behavior, without considering the user's reactions.

### 7.2.2 Measures

The dependent variables measured after the interaction with the ECA are as follows:

- User's perception of the agent's warmth (*w*) and competence (*c*): participants were asked to rate their level of agreement about how well each adjective described the ECA (4 adjectives concerning warmth and four concerning competence, according to Aragonés et al. (2015)). Even though only one dimension was manipulated at a time, we measured the user's impressions about both of them in order to check whether the manipulation of one dimension can affect the impressions about the other (as already found in the literature (Rosenberg et al., 1968; Judd et al., 2005; Yzerbyt, 2005)).
- User's experience of the interaction (*exp*): participants were asked to rate their level of agreement about a list of items adapted from the study by (Bickmore et al., 2011).

### 7.2.3 Hypotheses

We hypothesized the following scenarios:

- H1: when the ECA is in the Warmth condition, that is, when it adapts its behaviors according to the user's impressions of the agent's warmth, it will be perceived as warmer than it is in the Random condition;
- H2: when the ECA is in the Competence condition, that is, when it adapts its behaviors according to the user's impressions of the agent's competence, it will be perceived as more competent than it is in the Random condition;
- H3: when the agent ECA adapts its behaviors, that is, in either the Warmth or Competence conditions, this will improve the user's experience of the interaction, compared to that in the Random condition.

## 7.3 Analysis and Results

The visitors (24 women and 47 men) of the Carrefour Numérique of the Cité des sciences et de l'industrie of Paris were invited to take part in our experiment. 28% of them were in the range of 18–25 years old, 18% were in the range of 25–36, 28% were in the range of 36–45, 15% were in the range of 46–55, and 11% were over 55 years old. Participants were randomly assigned to each condition, with 25 participants assigned to the Warmth condition, 27 to the Competence condition, and 19 to the Random one.

We computed Cronbach's alphas on the scores of the four items about *w* and the four about *c*: good reliability was found for



**TABLE 1** | Mean and standard deviation of *w* and *c* scores for each level of Model.

Model	Warmth	Competence
Warmth	3.48 ± 0.8	3.2 ± 0.75
Competence	3.51 ± 0.96	3.3 ± 0.69
Random	3.26 ± 0.93	2.76 ± 0.73

both ( $\alpha = 0.85$  and  $\alpha = 0.81$ , respectively). Then, we computed the mean of these items in order to have one *w* score and one *c* score for each participant, and we used them for our analyses.

Since NARS scores got an acceptable degree of reliability ( $\alpha = 0.69$ ), we computed the overall mean of these items for each participant and divided them into two groups, “high” and “low,” according to whether they obtained a score higher than the overall mean or not, respectively. Participants were almost equally distributed into the two groups (35 in the “high” group and 36 in the “low” group). Chi-square tests for Model, age, and sex were run to verify that participants were equally distributed across these variables, too (all  $p > 0.5$ ).

### 7.3.1 Warmth Scores

The *w* means were normally distributed (the Shapiro test's  $p = 0.07$ ), and their variances were homogeneous (the Bartlett tests'  $ps$  for each variable were  $> 0.44$ ). We run a  $3 \times 5 \times 2 \times 2$  between-subjects ANOVA, with Model, age, sex, and NARS as factors.

No effects of age or sex were found. A main effect of NARS was found ( $F(1, 32) = 4.23, p < 0.05$ ). A *post hoc* test specified that the group who got high scores in NARS gave higher ratings about the agent's *w* ( $M = 3.65, SD = 0.84$ ) than the group who got low scores in NARS ( $M = 3.24, SD = 0.96$ ).

Although we did not find any significant effect, *w* scores were, on average, higher in the Warmth and Competence conditions than in the Random condition. The mean and standard error of *w* scores are shown in **Table 1**.

### 7.3.2 Competence Scores

The *c* means were normally distributed (the Shapiro test's  $p = 0.22$ ), and their variances were homogeneous (the Bartlett tests'  $ps$  for each variable were  $> 0.25$ ). We run a  $3 \times 5 \times 2 \times 2$  between-subjects ANOVA, with Model, age, sex, and NARS scores as factors.

We did not find any effect of age, sex, or NARS. A significant main effect of Model was found ( $F(2, 32) = 3.22, p = 0.047, \eta^2 = 0.085$ ). In particular, *post hoc* tests revealed that participants in the Competence condition gave higher scores about the agent's *c* than participants in the Random condition ( $M_C = 3.3, M_R = 2.76, p\text{-adj} = 0.05$ ).

### 7.3.3 User's Experience Scores

The *exp* items' means were not normally distributed, but their variances were homogeneous (the Bartlett tests'  $ps$  for each variable were  $> 0.17$ ). We run nonparametric tests for each item and each variable.

Even if we did not find any statistically significant effect, on average, items' scores tended to be higher in the Warmth and Competence conditions than in the Random condition.

### 7.3.4 Performance of the Adaptation Model

The Q-learning algorithm ended up selecting (for each participant) one specific combination of verbal and nonverbal behaviors from the  $84\% \pm 7$  and  $82\% \pm 7$  of the interaction, for the Warmth and Competence conditions, respectively. In the Warmth condition, the rest pose Akimbo was the most selected one ( $\chi^2 = 8.05, p < 0.01$ ), and we found a tendency to use Ideational gestures ( $p > 0.05$ ). In the Competence condition, the Verbal Behavior aiming at eliciting low warmth and high competence (formal language, long sentences, and use of you-words) was the most selected one ( $\chi^2 = 3.86, p < 0.01$ ).

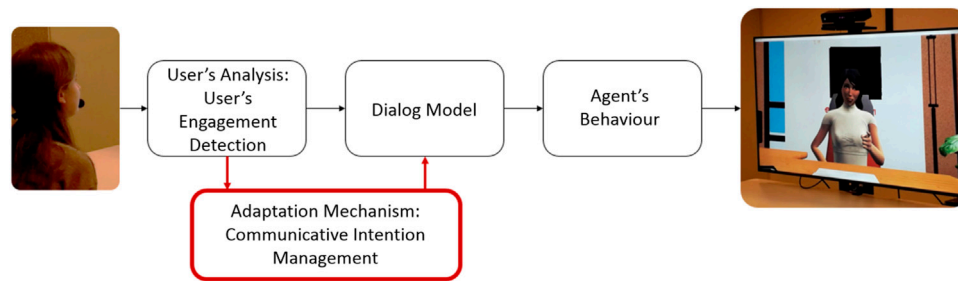
## 7.4 Discussion

The results show that participants' ratings tended to be higher in the conditions in which the ECA used the adaptation model than when it selected its behavior randomly. In particular, the results indicate that we successfully manipulated the impression of competence when using our adaptive ECA. Indeed, higher competence was reported in the Competence condition than in the Random one. No *a priori* effect was found.

On the other hand, we found an *a priori* effect on warmth but no significant effect of our conditions (just a positive trend for both the Competence and Warmth conditions). People with high *a priori* about virtual agents gave higher ratings about the agent's warmth than people with low *a priori*.

We could hypothesize some explanations for these results. First, we did not get the effects of our experimental conditions on warmth ratings since people were more anchored into their *a priori*, and it was hard to change them. Indeed, people's expectancies have already been found to have an effect on the user's judgments about ECAs (Burgoon et al., 2016; Biancardi et al., 2017b; Weber et al., 2018). The fact that we found this effect only for warmth judgments could be related to the primacy of warmth judgments over competence (Wojciszke and Abele, 2008). Then, it could have been easier to elicit impressions of competence since we found no *a priori* effect on competence. This could be explained as follows: people might expect that it is easier to implement knowledge in an ECA rather than social behaviors.

The user's experience of the interaction was not affected by the agent's adaptation. During the debriefing, many participants expressed their disappointment about the agent's appearance, the quality of the voice synthesizer and the animation, described as “disturbing” and “creepy,” and the limitations of the conversation (participants could only answer the ECA's questions). These factors could have reduced any other effect of the independent variables. Indeed, the agent's appearance and the structure of the dialogue were the same across conditions. If participants mainly focused on these elements, they could have paid less attention to the ECA's verbal and nonverbal behavior (the variables that were manipulated and that we were interested in), which thus did not manage to affect their overall experience of the interaction.



**FIGURE 4 |** Modified system architecture used in Study 2. In particular, the User's Analysis module contains the model to detect the user's engagement from facial and head/trunk signals. The Communicative Intention module uses reinforcement learning to select the agent's self-presentational strategy.

## 8 STUDY 2: ADAPTATION OF COMMUNICATIVE STRATEGIES

At this step, we investigate adaptation at a higher level than the previous one, namely, the communicative strategies of the ECA. In particular, we focus on the agent's self-presentational strategies, that is, different techniques to convey different levels of warmth and competence toward the user (Jones and Pittman, 1982). Each strategy is realized in terms of the verbal and nonverbal behavior of the ECA, according to the studies by (Pennebaker, 2011; Callejas et al., 2014; Biancardi et al., 2017a).

While in the previous study, we investigated whether and how adaptation could affect the user's impressions of the agent, we here focus on whether and how adaptation can affect the user's engagement during the interaction.

The goal of this second model is thus to make the ECA learn the communicative strategies that improve the user's engagement, by measuring and using the user's engagement as a reward.

### 8.1 Architecture

The general architecture described in Section 5 has been modified in order to contain a module for the detection of the user's engagement and a communicative intention planner for the choice of the agent's self-presentational strategy.

The modified architecture of the system is depicted in Figure 4. In the following subsection, we give more details about the modified modules.

#### 8.1.1 User's Analysis: User's Engagement Detection

The User's Analysis module is integrated with a User's Engagement Detection module that continuously computes the overall user's engagement at the end of every speaking turn. The computational model of the user's engagement is based on the detection of facial signals and head/trunk signals, which are indicators of engagement. In particular, smiling is usually considered an indicator of engagement, as it may show that the user is enjoying the interaction (Castellano et al., 2009). Eyebrows are equally important: for example, Corrigan et al. (2016) claimed that "frowning may indicate effortful processing suggesting high levels of cognitive engagement." Head/trunk signals are detected in order to measure the user's attention level. According to Corrigan et al. (2016), attention is

a key aspect of engagement; an engaged user continuously gazes at relevant objects/persons during the interaction. We approximate the user's gaze using the user's head and trunk orientation.

#### 8.1.2 Adaptation Mechanism: Communicative Intention Management

During its interaction with the user, the agent has the goal of selecting its self-presentational strategy (e.g., to communicate verbally and nonverbally a given dialogue act with high warmth and low competence). The agent can choose its strategy from a given set of four strategies inspired from Jones and Pittman's taxonomy (Jones and Pittman, 1982):

- **Ingratiation:** the ECA has the goal to convey positive interpersonal qualities and elicit impressions of high warmth toward the user, without considering its level of competence;
- **Supplication:** the ECA has the goal to present its weaknesses and elicit impressions of high warmth and low competence;
- **Self-promotion:** the ECA has the goal to focus on its capabilities and elicit impressions of high competence, without considering its level of warmth; and
- **Intimidation:** the ECA has the goal to elicit impressions of high competence by decreasing its level of warmth.

The verbal behavior characterizing the different strategies is inspired by the works of Pennebaker (2011) and Callejas et al. (2014). In particular, we took into account the use of you- and we-words, the level of formality of the language, and the length of the sentences.

The choice of the agent's nonverbal behavior is based on our previous studies (Biancardi et al., 2017a; Biancardi et al., 2017b). So, for example, if the current agent's self-presentational strategy is Supplication and the next dialogue act to be spoken is introducing a topic, then the agent would say "I think that while you play there are captors that measure tons of stuffs!" accompanied by smiling and beat gestures. Conversely, if the current agent's self-presentational strategy is Intimidation and the next dialogue act to be spoken is the same, then the agent would say "While you play at video games, several captors

**TABLE 2 |** Mean and standard deviation values of warmth scores for each level of Communicative Strategy. The mean score for Intim\_static is significantly lower than that for all the other conditions.

Communicative Strategy	Warmth
Ingr_static	3.77 ± 0.57
Supp_static	3.54 ± 0.999
Self_static	3.81 ± 0.70
Intim_static	2.63 ± 0.93
Random	3.71 ± 0.80
Adaptation	3.89 ± 0.38

measure your physiological signals,” accompanied by ideational gestures without smiling.

To be able to change the agent’s communicative strategy according to the detected participant’s engagement, we applied a reinforcement learning algorithm to make the ECA learn what strategy to use. Specifically, a multiarmed bandit algorithm (Katehakis and Veinott, 1987) was applied. This algorithm is a simplified setting of reinforcement learning which models agents evolving in an environment where they can perform several actions, each action being more or less rewarding for them. The choice of the action does not affect the state (i.e., what happens in the environment). In our case, the actions that the ECA could perform are the verbal and nonverbal behaviors corresponding to the self-presentational strategy that the ECA aims to communicate. The environment is the interaction with the user, while the state space is the set of dialogue acts used at each speaking turn. The choice of the action does not change the state (i.e., the dialogue act used during the actual speaking turn), but rather, it acts on how this dialogue act is realized by verbal and nonverbal behavior. More details about the multiarmed bandit function used in our model can be found in the study by (Biancardi et al., 2019a).

## 8.2 Experimental Design

The adaptation model described in Section 8.1.2 was evaluated by using the scenario described in Section 6. Here, we describe the experimental variables manipulated and measured during the experiment.

### 8.2.1 Independent Variable

The design includes one independent variable, called Communicative Strategy, with six levels determining the way in which the ECA chooses the strategy to use:

- 1) Adaptation: the ECA uses the adaptation model and thus selects one self-presentational strategy at each speaking turn, by using the user’s engagement as a reward;
- 2) Random: the ECA chooses a random behavior at each speaking turn;
- 3) Ingr\_static: the ECA always adopts the Ingratiation strategy during the whole interaction;
- 4) Suppl\_static: the ECA always adopts the Supplication strategy during the whole interaction;

**TABLE 3 |** Mean and standard deviation values of competence scores for each level of Communicative Strategy. No significant differences among the conditions were found.

Communicative Strategy	Competence
Ingr_static	3.6 ± 0.62
Supp_static	2.98 ± 0.77
Self_static	3.75 ± 0.63
Intim_static	3.65 ± 0.79
Random	3.5 ± 0.70
Adaptation	3.43 ± 0.76

- 5) Self\_static: the ECA always adopts the Self-promotion strategy during the whole interaction; and
- 6) Intim\_static: the ECA always adopts the Intimidation strategy during the whole interaction.

### 8.2.2 Measures

The dependent variables measured after the interaction with the ECA are the same as those described in subsection 7.2.2.

In addition to these measures, during the interaction, for people who agreed with audio recording of the experiment, we collected quantitative information about their verbal engagement, in particular, the polarity of the user’s answer when the ECA asked if they wanted to continue to discuss and the number of any verbal feedback produced by the user during a speaking turn.

### 8.2.3 Hypotheses

We hypothesized that each self-presentational strategy would elicit the right degree of warmth and competence, in particular, the following:

H1ingr: the ECA in the Ingr\_static condition would be perceived as warm by users;

H1supp: the ECA in the Suppl\_static condition would be perceived as warm and not competent by users;

H1self: the ECA in the Self\_static condition would be perceived as competent by users; and

H1intim: the ECA in the Intim\_static condition would be perceived as competent and not warm by users.

Then, we hypothesized the following scenarios:

H2a: an ECA adapting its self-presentational strategies according to the user’s engagement would improve the user’s experience, compared to a non-adapting ECA and H2b: the ECA in the Adaptation condition would influence how it is perceived in terms of warmth and competence.

## 8.3 Analysis and Results

75 participants (30 females) took part in the evaluation, equally distributed among the six conditions. The majority of them were in the 18–25 or 36–45 age range and were native French speakers. In this section, we briefly report the main results of our analyses.

A more detailed report can be found in the study by (Biancardi et al., 2019a).

### 8.3.1 Warmth Scores

A  $4 \times 2$  between-subjects ANOVA revealed a main effect of Communicative Strategy ( $F(5,62) = 4.75$ ,  $p < 0.001$ ,  $\eta^2 = 0.26$ ) and NARS ( $F(1,62) = 5.74$ ,  $p < 0.05$ ,  $\eta^2 = 0.06$ ). The  $w$  ratings were higher from participants with a high NARS score ( $M = 3.74$ ,  $SD = 0.77$ ) than from those with a low NARS score ( $M = 3.33$ ,  $SD = 0.92$ ).

**Table 2** shows the mean and SD of  $w$  scores for each level of Communicative Strategy. Multiple comparisons t-test using Holm's correction shows that the  $w$  mean for Intim\_static is significantly lower than that for all the others. As a consequence, the other conditions are rated as warmer than Intim\_static. H1ingr and H1supp are thus validated, and H1intim and H2b are validated for the warmth component.

### 8.3.2 Competence Scores

No significant results emerged from the analyses. When looking at the means of  $c$  for each condition (see **Table 3**), Supp\_static is the one with the lower score, even if its difference with the other scores does not reach statistical significance (all  $p$ -values  $> 0.1$ ). H1supp and H1intim (for the competence component) are not validated.

### 8.3.3 User's Experience of the Interaction

Participants in the Ingr\_static condition were more satisfied from the interaction than those in Suppl\_static ( $z = 2.88$ ,  $p\text{-adj} < 0.05$ ) and in Intim\_static ( $z = 2.56$ ,  $p\text{-adj} < 0.05$ ). Participants in the Ingr\_static condition also liked the ECA more than participants in the Intim\_static condition ( $z = 2.87$ ,  $p\text{-adj} < 0.05$ ). No differences were found between the scores of the participants in the Adaptation condition and those of the other participants for any of the items measuring *exp*.

The *exp* scores are also affected by participants' *a priori* about virtual agents (measured through the NARS questionnaire). In particular, participants who got high scores in the NARS questionnaire were more satisfied by the interaction ( $U = 910.5$ ,  $p < 0.05$ ), were more motivated to continue the interaction ( $U = 998$ ,  $p = 0.001$ ), and perceived the agent as less closed to a computer ( $U = 1028$ ,  $p < 0.001$ ) than people who got low scores in the NARS questionnaire.

Another interesting result concerns the effect of age on participants' satisfaction ( $H(4) = 15.05$ ,  $p < 0.01$ ); people in the age range of 55+ were more satisfied than people of any other age range (all  $p\text{-adj} < 0.05$ ).

On the whole, these results do not allow us to validate H2a, but the agent's adaptation was found to have at least an effect on its level of warmth (H2b).

### 8.3.4 Verbal Cues of Engagement

During each speaking turn, the user was free to reply to the agent's utterances. We consider as a user's verbal feedback any type of verbal reply to the ECA, from a simple backchannel (e.g., "ok" and "mm") to a longer response (e.g., giving an opinion about what the ECA said). In general, participants who did not give much verbal feedback (i.e., less

than 13 replies to the agent's utterances over all the speaking turns) answered positively to the ECA when it asked whether they wanted to continue to discuss with it, compared to the participants who gave more verbal feedback ( $OR = 4.27$ ,  $p < 0.05$ ). In addition, we found that the participants who did not give much verbal feedback liked the ECA more than those who talked a lot during the interaction ( $U = 36.5$ ,  $p < 0.05$ ). However, no differences in any of the dependent variables were found according to Communicative Strategy.

## 8.4 Discussion

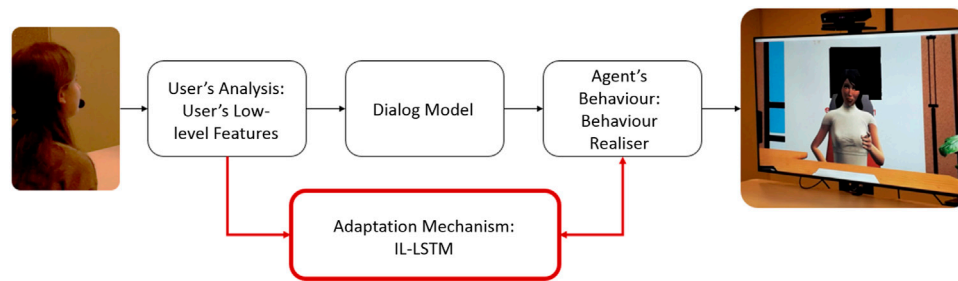
First of all, regarding H1, the only statistically significant results concern the perception of the agent's warmth. The ECA was rated as colder when it adopted the Intim\_static strategy than when it adopted the other conditions. This supports the thesis of the primacy of the warmth dimension (Wojciszke and Abele, 2008), and it is in line with the positive-negative asymmetry effect described by (Peeters and Czapinski, 1990), who argued that negative information generally has a higher impact on person perception than positive information. In our case, when the ECA displayed cold (i.e., low warmth) behaviors (i.e., in the Intim\_static condition), it was judged by participants with statistically significant lower ratings of warmth. Regarding the other conditions (Ingr\_static, Supp\_static, Self\_static, Adaptation, and Random), they elicited warmer impressions in the user, but there was not one strategy that was better than the others in this regard. The fact that Self\_static also elicited the same level of warmth as the others reflected a halo effect (Rosenberg et al., 1968); the behaviors displayed to appear competent influenced its warmth perception in the same direction.

Regarding H2, the results do not validate our hypothesis (H2a) that the interaction would be improved when the ECA managed its impressions by adapting its strategy according to the user's engagement. When analyzing scores for *exp* items, we found that participants were more satisfied by the interaction and they liked the ECA more when the ECA wanted to be perceived as warm (i.e., in the Ingr\_static condition) than when it wanted to be perceived as cold and competent (i.e., in the Intim\_static condition). A hypothesis is that since the ECA was perceived warmer in the Ingr\_static condition, it could have positively influenced the ratings of the other items, like the user's satisfaction. Concerning H2b with regard to a possible effect of the agent's adaptation on the user's perception of its warmth and competence, it is interesting to see that when the ECA adapted its self-presentational strategy according to the user's overall engagement, it was perceived as warm. This highlights a link between the agent's adaptation, the user's engagement, and a warm impression; the more the ECA adapted its behaviors, the more the user was engaged and the more she/he perceived the ECA as warm.

## 9 STUDY 3: ADAPTATION AT A SIGNAL LEVEL

At this step, we are interested in low-level adaptation at the signal level. We aim to model how the ECA can adapt its signals to the user's signals. Thus, we make the ECA predict the signals to





**FIGURE 5 |** Modified system architecture used in Study 3. In particular, the User's Analysis module detects the user's low-level signals such as head and eye rotations and lip corner activity. The Adaptation Mechanism module exploits the IL-LSTM model for selecting the agent's low-level signals. In the Agent's Behavior module, the Behavior Realizer is customized in order to take into account the agent's communicative behaviors and signals coming from the IL-LSTM module in real time.

display at each time step, according to those displayed by both the ECA and the user during a given time window. For the sake of simplicity, we consider a subset of signals, namely, lip corner movement (AU12), gaze direction, and head movement. To reach our aim, we follow a two-step approach. At first, we need to predict which signals that are due to adaptation to the user's behaviors should be displayed by the ECA at each time step. The prediction of signal adaptation is learned on human-human interaction. The ECA ought to communicate its intentions to adapt to the user's signals. Then, the second step of our approach consists in blending the predicted signals linked to the adaptation mechanism with the nonverbal behaviors corresponding to the agent's communicative intentions. We describe our algorithm in further detail in **subsection 9.1.2**.

## 9.1 Architecture

The general architecture described in **Section 5** has been modified in order to contain a module for predicting the next social signal to be merged with the agent's other communicative ones. The modified architecture of the system is depicted in **Figure 5**. In the following subsection, we explain the modified modules. More details about these modules can be found in the study by (Dermouche and Pelachaud, 2019).

### 9.1.1 User's Analysis: User's Low-Level Features

Low-level features of the user are obtained from the User's Analysis module using EyesWeb of the general architecture. In this model, we consider a subset of these features, namely, the user's head direction, eye direction, and AU12 (upper lip corner activity). At every frame, the EyesWeb module extracts these features and sends the last 20 analyzed frames to the Adaptation Mechanism module IL-LSTM (see **Section 9.1.2**). It also sends the user's conversational state (speaking or not) computed from the detection of the user's voice activity (done in EyesWeb) and from the agent-turn information provided by the dialogue manager Flipper.

### 9.1.2 Adaptation Mechanism: Interaction Loop-LSTM

In this version of the architecture, the adaptation mechanism is based on a predictive model trained on data of human-human interactions. We used the NoXi database (Cafaro et al., 2017) to train

a long short-term memory (LSTM) model that takes as input sequences of signals of two interactants over a sliding window of  $n$  frames to predict which signal(s) should display one participant at time  $n+1$ . We call this model IL-LSTM, which stands for interaction loop-LSTM. LSTM is a kind of recurrent neural network. It is mainly used when "context" is important, that is, decisions from the past can influence the current ones. It allows us to model both sequentiality and temporality of nonverbal behaviors.

We apply the IL-LSTM model to the human-agent interaction. Thus, given the signals produced by both, the human and the ECA, over a time window, the model outputs which signals should display the ECA at the next time step (here, a frame). The predicted signals are sent to the Behavior Realizer of the Agent's Behavior module where they are merged with the behaviors of the ECA related to its communicative intents.

### 9.1.3 Agent's Behavior: Behavior Realizer

We have updated the Behavior Realizer so that the ECA not only communicates its intentions but also adapts its behaviors in real time to the user's behaviors. This module blends the predicted signals linked to the adaptation mechanism with the nonverbal behaviors corresponding to its communicative intentions that have been outputted using the GRETA agent platform (Pecune et al., 2014). More precisely, the dialogue module Flipper sends the set of communicative intentions to the Agent's Behavior module. This module computes the multimodal behavior of the ECA and sends it to the Behavior Realizer that computes the animation of the ECA's face and body. Then, before sending each frame to be displayed by the animation player, the animation computed from the communicative intentions is merged with the animation predicted by the Adaptation Mechanism module. This operation is repeated at every frame.

## 9.2 Experimental Design

The adaptation model described in the previous section was evaluated by using the scenario described in **Section 6**. Here, we describe the experimental variables manipulated and measured during the experiment.

### 9.2.1 Independent Variable

We manipulated the type of low-level adaptation of the ECA by considering five conditions:

**TABLE 4 |** Mean  $\pm$  standard deviation of each dimension of the questionnaires (each row of the table), for each of the five conditions (each column).

	Random	Head	Lip Corners	Eyes	All
Competence	2.98 $\pm$ 1.22	3.45 $\pm$ 0.81	3.73 $\pm$ 0.73	3.65 $\pm$ 1.06	3.61 $\pm$ 1.12
Distance	2.5 $\pm$ 1.12	2.6 $\pm$ 1.03	1.76 $\pm$ 0.97	2 $\pm$ 1.12	1.47 $\pm$ 1.03
Friendliness	3.03 $\pm$ 1.12	3.22 $\pm$ 0.86	3.80 $\pm$ 0.83	3.33 $\pm$ 0.86	4.09 $\pm$ 0.90 <sup>a</sup>
Involvement	2.65 $\pm$ 1.22	2.65 $\pm$ 1.15	3.52 $\pm$ 1.00 <sup>a</sup>	2.83 $\pm$ 1.33	3.60 $\pm$ 1.07
Realism	1.7 $\pm$ 0.92	1.95 $\pm$ 0.82	2.52 $\pm$ 1.12	2.08 $\pm$ 0.90	1.73 $\pm$ 0.86
Relevance	2.95 $\pm$ 1.38	3.86 $\pm$ 0.72	3.97 $\pm$ 0.79 <sup>a</sup>	3.5 $\pm$ 1.24	3.80 $\pm$ 1.01
Satisfaction	2.46 $\pm$ 1.21	2.84 $\pm$ 0.77	3.39 $\pm$ 0.06	3.27 $\pm$ 1.08	3.39 $\pm$ 0.93

<sup>a</sup>indicates that the score is significantly different compared to that in the Random condition ( $p - \text{adj} < .05$ ).

- Random: when the ECA did not adapt its behavior;
- Head: when the ECA adapted its head rotation according to the user's behavior;
- Lip Corners: when the ECA adapted its lip corner puller movement (AU12) according to the user's behavior;
- Eyes: when the ECA adapted its eye rotation according to the user's behavior; and
- All: when the ECA adapted its head and eye rotation and lip corner movement, according to the user's behavior.

We tested these five conditions using a between-subjects design.

### 9.2.2 Measures

The dependent variables measured after the interaction with the ECA were the user's engagement and the perceived friendliness of the ECA.

The user's engagement was evaluated using the I-PEFiC framework (van Vugt et al., 2006) that encompasses the user's engagement and satisfaction during human-agent interaction. This framework considers different dimensions regarding the perception of the ECA (in terms of realism, competence, and relevance) as well as the user's engagement (involvement and distance) and the user's satisfaction. We adapted the questionnaire proposed by Van Vugt and others to measure the behavior of the ECA along these dimensions (van Vugt et al., 2006). The perceived friendliness of the ECA was measured using the adjectives kind, warm, agreeable, and sympathetic of the IAS questionnaire (Wiggins, 1979).

As for the other two studies, we also measured the *a priori* attitude of participants towards virtual agents using the NARS questionnaire.

### 9.2.3 Hypotheses

Previous studies (Liu et al., 2008; Woolf et al., 2009; Levitan, 2013) have found that users' satisfaction about their interaction with an ECA is greater when the ECA adapts its behavior to the user's one. From these results, we could expect that the user would be more satisfied about the interaction when the ECA adapted its low-level signals according to their behaviors. We also assumed that the ECA adapting its lip corner puller (that is related to smiling) would be perceived as friendlier. Thus, our hypotheses were as follows:

H1Head: when the ECA adapted its head rotation, the users would be more satisfied with the interaction than the users interacting with the ECA in the Random condition.

H2aLips: when the ECA adapted its lip corner movement (AU12), the users would be more satisfied with the interaction than the users interacting with the ECA in the Random condition.

H2bLips: when the ECA adapted its lip corner movement (AU12), it would be evaluated as friendlier than the ECA in the Random condition.

H3Eyes: when the ECA adapted its eye rotation, the users would be more satisfied with the interaction than the users interacting with the ECA in the Random condition.

H4aAll: when the ECA adapted its head and eye rotations and lip corner movement, the users would be more satisfied with the interaction than the users interacting with the ECA in the Random condition.

H4bAll: when the ECA adapts its head and eye rotations and lip corner movement, it would be evaluated as friendlier than the ECA in the Random condition.

## 9.3 Analysis and Results

101 participants (55 females), almost equally distributed among the five conditions, took part in our experiment. 95% of participants were native French speakers. 32% of them were in the range of 18–25 years old, 17% were in the range of 25–36, 21% were in the range of 36–45, 18% were in the range of 46–55, and 12% were over 55 years old. For each dimension of the user's engagement questionnaire, as well as for that about the perceived friendliness of the ECA, Cronbach's  $\alpha$ s were  $> 0.8$ ; we then computed the mean of the scores in order to have one score for each dimension. The mean and standard deviation of each measured dimension for each of the five conditions are shown in Table 4.

As our data were not normally distributed (the Shapiro test's  $p < 0.5$ ), we used the unpaired Wilcoxon test (equivalent to *t*-test) to measure how participants' ratings differed between the Random condition and each of the other conditions.

In the Head condition, we could not find differences with the Random condition. We conclude that the hypothesis H1Head is rejected.

In the Lip Corners condition, compared to participants in the Random condition, participants were more involved ( $W = 98.5$ ,  $p - \text{adj} < .05$ ). We can also note that the ECA was evaluated as more positive on the relevance dimension ( $W = 104.5$ ,  $p - \text{adj} < .05$ ). We can conclude that the hypotheses H2aLips and H2bLips are not validated, but the adaptation of lip corner movement still has a positive effect on other dimensions related to the user's engagement.

In the Eyes condition, participants were satisfied with the ECA as they were with the ECA in the Random condition. Thus, the hypothesis H3Eyes is rejected.

In the All condition, the ECA was evaluated as friendlier ( $W = 104.5$ ,  $p\text{-adj} < .05$ ) than the ECA in the Random condition. So, H4aAll is supported, while H4bAll is rejected.

Results of the NARS questionnaire indicated that 40, 30, and 30% of participants, respectively, had a positive, neutral, and negative attitude toward virtual agents. An ANOVA test was performed to study the influence of participants' *a priori* toward virtual agents on their engagement in the interaction. Participants' prior attitude toward ECAs had a main effect on participants' distance ( $F(1, 93) = 5.13$ ,  $p < .05$ ). Results of pairwise comparisons with Bonferroni adjustment highlighted that participants with a prior negative attitude were less engaged (more distant ( $p\text{-adj} < .05$ ) and less involved ( $p\text{-adj} < .05$ )) than those with a prior positive attitude.

## 9.4 Discussion

The results of this study showed that participants' engagement and perception of the ECA's friendliness were positively impacted when the ECA adapted its low-level signals.

These results were significant only when the ECA adapted its lip corner movement (AU12) to the user's behavior (mainly their smile), that is, in the Lip Corners and All conditions. In the case of head and eye rotation adaptation, we found a trend on some dimensions but no significant differences compared to the Random condition. These results could be caused by the adopted evaluation setting where the ECA and the user faced each other. During the interaction, most participants gazed at the ECA without doing any postural shift or even changing their gaze and head direction. They were mainly still and staring at the ECA. The adaptive behaviors, that is, head and eye rotation of the ECA computed from the user's behaviors, remained constant throughout the interaction. They reflected participants' behaviors (that were not moving much). Thus, in the Head and Eyes adapting conditions, the ECA showed much less expressiveness and may have appeared much less lively, which may have impacted participants' engagement in the interaction.

## 10 GENERAL DISCUSSION

In our studies, we applied the interaction adaptation theory (see Section 2) on the ECA. That is, our adapting ECA had the requirement R that it needed to adapt in order to have a successful interaction. Its desire D was to maximize the user's experience by eliciting a specific impression toward the user or maintaining the user's engagement. Finally, its expectations (Es) were that the user's experience would be better when interacting with an adaptive ECA. All these factors rely on the general hypothesis that the user expects to interact with a social entity. According to this hypothesis, the ECA should adapt its behavior like humans do (Appel et al., 2012).

We have looked at different adaptation mechanisms through three studies, each focusing on a specific type of adaptation. In our studies, we found that these mechanisms impacted the user's experience of the interaction and their perception of the ECA. Moreover, in all three studies, interacting with an adaptive ECA

vs. a nonadaptive ECA tended to be more positively perceived. More precisely, manipulating the agent's behaviors (Study 1) had an impact on the user's perception of the ECA while low-level adaptation (Study 3) positively influenced the user's experience of the interaction. Regarding managing conversational strategies (Study 2), the ECA was perceived as warmer when it managed those that increased the user's engagement vs. when it did not change them all along the interaction.

These results suggest that the IAT framework allows for enhancing human-agent interaction. Indeed, the adaptive ECA shows some improvement in the quality of the interaction and the perception of the ECA in terms of social attitudes.

However, not all our hypotheses were verified. This could be related to the fact that we based our framework on the general hypothesis that the user expects to interact with a social entity. The ECA did not take into account the fact that the user also had their specific requirements, desires, and expectations, along with the expectancy to interact with a social agent. Yet, the ECA did not check if the user still considered it a social entity during the interaction. It based its behaviors only on the human's detected engagement and impressions. Moreover, the modules to detect engagement or impressions work in a given time window, but they do not consider their evolution through time. For example, the engagement module computes that participants are engaged if they look straight at the ECA without reporting any information stating that the participants stare fixedly at the ECA. The fact that participants do not change their gaze direction toward the ECA could be interpreted as participants not viewing the ECA as a social entity with humanlike qualities (Appel et al., 2012).

Expectancy violation theory (Burgoon, 1993) could help to better understand this gap. This theory explains how confirmations and violations of people's expectancies affect communication outcomes such as attraction, liking, credibility persuasion, and learning. In particular, positive violations are predicted to produce better outcomes than positive confirmations, and negative violations are predicted to produce worse outcomes than negative confirmations. Expectancy violation theory has already been demonstrated to affect human-human interaction (Burgoon, 1993) and when people are in front of an ECA (Burgoon et al., 2016; Biancardi et al., 2017b) or a robot (Weber et al., 2018). In our work, we took into account the role of expectancies as part of IAT. Our results suggest that expectancies could play a more important role than the one we attributed to them and that they should be better modeled when developing human-agent adaptation. Future works in this context should combine expectancy violation theory with IAT. In this way, the ECA should be able to detect the user's expectancies in terms of beliefs and desires. It should also be able to check if those expectancies about the interaction correspond to the expected ones and then react accordingly. For example, in our studies, we found some effects of people's *a priori* about virtual agents: people who got higher scores in the NARS questionnaire generally perceived the ECA as warmer than people who got lower scores in the NARS questionnaire. This effect could have been mitigated if the agent could detect the user's *a priori*.

Even with these limits, the results of our studies show that an adaptive model for a virtual agent inspired from IAT partially managed to produce an impact on the user's experience of the interaction and on their perception of the ECA. This could be useful for personalizing systems for different applications such as education, healthcare, or

entertainment, where there is a need of adaptation according to users' type and behaviors and/or interaction contexts.

The different adaptation models we developed also confirm the potential of automatic behavior analysis for the estimation of different users' characteristics. These methods can be used to better understand the user's profile and can also be applied to human-computer interaction in general to inform adaptation models in real time.

Moreover, the use of adaptation mechanisms inspired from IAT could help mitigate the negative effect of some interaction problems that are more difficult to solve, due to, for example, technological limits of the system. Indeed, adaptation acts to enhance the agent's perception and the perceived interaction quality. Improving adaptation mechanisms may help to counterbalance technological shortcomings. It may also improve the acceptability of innovative technologies that are likely to be part of our daily lives, in the context of work, health, leisure, etc.

## 11 CONCLUSION AND FUTURE WORK

In this study, we investigated adaptation in human-agent interaction. In particular, we reported our work about three models focusing on different levels of the agent's adaptation (the behavioral, conversational, and signal levels), by framing them in the same theoretical framework (Burgoon et al., 2007). In all the adaptation mechanisms implemented in the models, the user's behavior is taken into account by the ECA during the interaction in real time. Evaluation studies showed a tendency toward a positive impact of the adaptive ECA on the user's experience and perception of the ECA, encouraging us to continue to investigate in this direction.

One limitation of our models is their reliance on the interaction scenario. Indeed, to obtain good performances of adaptation models using reinforcement learning algorithms, a scenario including an adequate number of steps is required. In our case, the agent ended up selecting a specific combination of behaviors only during the later part of the interaction. A longer interaction with more steps would allow an adaptive agent using reinforcement learning algorithms to better learn. Another possibility would be to have participants interacting more than once with the virtual agent. This latter would require adding a memory adaptation module (Ahmad et al., 2017). This would also allow for checking whether the same user prefers the same behavior and/or conversational strategies from the agent over several interactions. Similarly, regarding adaptation models reflecting the user's behavior, the less the user moves during the interaction, the less the agent's expressivity level. The interaction scenario should be designed in order to elicit the user's participation, including strategies to tickle users when they become too still and nonreactive. For example, one could use a scenario including a collaborative task where both the agent and the user would interact with different objects. In such a setting, although it would require us to extend our engagement detection module to include joint attention, we expect that the participants would also perform many more head movements that, in turn, could be useful for a better low-level adaptation of the agent.

In the future, our work could be improved and explored along further axes. We list three of them here. First, the three models

presented in this article were implemented and evaluated independently from each other. It could be interesting to merge the three adaptation mechanisms in a broader model and investigate the impacts of the agent's adaptation along different levels at the same time. Second, in our studies, the agent adapted its behaviors to the user's ones without considering if the relationship between the behaviors of the dyad showed any specific interaction patterns. In particular, we have not made explicit if the agent's behavior should either match, reciprocate, complement, compensate, or mirror their human interlocutor's behavior (Burgoon et al., 2007). Also, we have not measured any similarities, synchronization, or imitation between the user's and the agent's behavior when we analyzed the data of our studies. Since adaptation may be signaled through a larger variety of behavior manifestations during an interaction, more adaptation mechanisms could be implemented. One last important direction for future work concerns the improvement of the interaction with the user. This would reduce possible secondary effects of uncontrolled variables, such as the user's expectancies, and allow for better studying of the effects of the agent's adaptation. We aim to improve the agent's conversational skills to ensure conversation repairs and interruptions and by letting the user choose the topic of conversation (e.g., from a set of possible ones) and drive the discussion. In addition to these improvements, the user's expectancies should also be better modeled by taking into account expectancy violation theory in addition to interaction adaptation theory.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

## ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

BB and SD: conceptualization, methodology, investigation, formal analysis, and writing—original draft preparation. CP: writing—review and editing, supervision, and funding acquisition.

## FUNDING

The research topics addressed in this article have been investigated in the framework of the EU Horizon 2020 research and innovation program under Grant Agreement Number 769553 and the ANR project Impressions ANR-15-CE23-0023.



## REFERENCES

- Ahmad, M. I., Mubin, O., and Orlando, J. (2017). Adaptive Social Robot for Sustaining Social Engagement during Long-Term Children-Robot Interaction. *Int. J. Human-Computer Interaction* 33 (12), 943–962. doi:10.1080/10447318.2017.1300750
- Altman, I., Vinsel, A., and Brown, B. B. (1981). Dialectic Conceptions in Social Psychology: An Application to Social Penetration and Privacy Regulation. *Adv. Exp. Soc. Psychol.* 14, 107–160. doi:10.1016/s0065-2601(08)60371-8
- Andersen, P. A. (1985). “Nonverbal Immediacy in Interpersonal Communication,” in *Multichannel Integrations of Nonverbal Behavior*. Editors A. A. Siegman and S. Feldstein (Hillsdale, NJ: Erlbaum: Psychology Press), 1–36.
- Appel, J., von der Pütten, A., Krämer, N. C., and Gratch, J. (2012). Does Humanity Matter? Analyzing the Importance of Social Cues and Perceived agency of a Computer System for the Emergence of Social Reactions during Human-Computer Interaction. *Adv. Human-Computer Interaction* 2012, 324694. doi:10.1155/2012/324694
- Aragónés, J. I., Poggio, L., Seviliano, V., Pérez-López, R., and Sánchez-Bernardos, M.-L. (2015). Measuring warmth and competence at inter-group, interpersonal and individual levels/Medición de la cordialidad y la competencia en los niveles intergrupales, interindividual e individual. *Revista de Psicología Soc.* 30 (3), 407–438. doi:10.1080/02134748.2015.1065084
- Argyle, M., and Dean, J. (1965). Eye-contact, Distance and Affiliation. *Sociometry* 28 (3), 289–304. doi:10.2307/2786027
- Argyle, M. (1972). *Non-verbal Communication in Human Social Interaction*. New York: Cambridge University Press.
- Bailenson, J. N., and Yee, N. (2005). Digital Chameleons: Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments. *Psychol. Sci.* 16, 814–819. doi:10.1111/j.1467-9280.2005.01619.x
- Baltrušaitis, T., Robinson, P., and Morency, L.-P. (2016). “Openface: an Open Source Facial Behavior Analysis Toolkit,” in Applications of Computer Vision (WACV), 2016 IEEE Winter Conference, Lake Placid, NY, USA, 7–10 March 2016 (IEEE), 1–10. doi:10.1109/WACV.2016.7477553
- Bergmann, K., Eyssel, F., and Kopp, S. (2012). “A Second Chance to Make a First Impression? How Appearance and Nonverbal Behavior Affect Perceived Warmth and Competence of Virtual Agents over Time,” in International Conference on Intelligent Virtual Agents, Santa Cruz, CA, USA, September, 12–14 (Springer), 126–138. doi:10.1007/978-3-642-33197-8\_13
- Bernieri, F. J., Reznick, J. S., and Rosenthal, R. (1988). Synchrony, Pseudosynchrony, and Dissynchrony: Measuring the Entrainment Process in Mother-Infant Interactions. *J. Personal. Soc. Psychol.* 54 (2), 243–253. doi:10.1037/0022-3514.54.2.243
- Bevacqua, E., Mancini, M., and Pelachaud, C. (2008). “A Listening Agent Exhibiting Variable Behavior,” in International Conference on Intelligent Virtual Agents, Tokyo, Japan, September 1–3, 2008 (Springer), 262–269. doi:10.1007/978-3-540-85483-8\_27
- Biancardi, B., Cafaro, A., and Pelachaud, C. (2017a). “Analyzing First Impressions of Warmth and Competence from Observable Nonverbal Cues in Expert-Novice Interactions,” in Proceedings of the 19th ACM International Conference on Multimodal Interaction, November 2017 (Glasgow: ACM), 341–349. doi:10.1145/3136755.3136779
- Biancardi, B., Cafaro, A., and Pelachaud, C. (2017b). “Could a Virtual Agent Be Warm and Competent? Investigating User’s Impressions of Agent’s Non-verbal Behaviors,” in Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents, November 2017 (Glasgow: ACM), 22–24. doi:10.1145/3139491.3139498
- Biancardi, B., Mancini, M., Lerner, P., and Pelachaud, C. (2019a). Managing an Agent’s Self-Presentational Strategies during an Interaction. *Front. Robot. AI* 6, 93. doi:10.3389/frobt.2019.00093
- Biancardi, B., Wang, C., Mancini, M., Cafaro, A., Chanel, G., and Pelachaud, C. (2019b). “A Computational Model for Managing Impressions of an Embodied Conversational Agent in Real-Time,” in 2019 International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, UK, 3–6 Sept. 2019 (IEEE). doi:10.1109/ACII.2019.8925495
- Bickmore, T., Pfeifer, L., and Schulman, D. (2011). “Relational Agents Improve Engagement and Learning in Science Museum Visitors,” in International Conference on Intelligent Virtual Agents, Reykjavik, Iceland, September 15–17, 2011 (Springer), 55–67. doi:10.1007/978-3-642-23974-8\_7
- Bickmore, T. W., Vardoulakis, L. M. P., and Schulman, D. (2013). Tinker: a Relational Agent Museum Guide. *Auton. Agent Multi-agent Syst.* 27 (2), 254–276. doi:10.1007/s10458-012-9216-7
- Burgoon, J. K., Bonito, J. A., Lowry, P. B., Humpherys, S. L., Moody, G. D., Gaskin, J. E., et al. (2016). Application of Expectancy Violations Theory to Communication with and Judgments about Embodied Agents during a Decision-Making Task. *Int. J. Human-Computer Stud.* 91, 24–36. doi:10.1016/j.ijhcs.2016.02.002
- Burgoon, J. K. (1993). Interpersonal Expectations, Expectancy Violations, and Emotional Communication. *J. Lang. Soc. Psychol.* 12 (1–2), 30–48. doi:10.1177/0261927x93121003
- Burgoon, J. K., Stern, L. A., and Dillman, L. (2007). *Interpersonal Adaptation: Dyadic Interaction Patterns*. New York: Cambridge University Press.
- Cafaro, A., Vilhjálmsson, H. H., and Bickmore, T. (2016). First Impressions in Human-Agent Virtual Encounters. *ACM Trans. Comput.-Hum. Interact.* 23 (4), 1–40. doi:10.1145/2940325
- Cafaro, A., Wagner, J., Baur, T., Dermouche, S., Torres Torres, M., Pelachaud, C., André, E., and Valstar, M. (2017). “The NoXi Database: Multimodal Recordings of Mediated Novice-Expert Interactions,” in Proceedings of the 19th ACM International Conference on Multimodal Interaction, November 2017 (Glasgow: ACM), 350–359. doi:10.1145/3136755.3136780
- Callejas, Z., Ravenet, B., Ochs, M., and Pelachaud, C. (2014). “A Computational Model of Social Attitudes for a Virtual Recruiter,” in Proceedings of the 13th international conference on Autonomous Agents and Multi-Agent Systems, May 2014 (Paris: ACM), 93–100.
- Camurri, A., Coletta, P., Massari, A., Mazzarino, B., Peri, M., Ricchetti, M., Ricci, A., and Volpe, G. (2004). “Toward Real-Time Multimodal Processing: Eyesweb 4.0,” in Proceedings of the Artificial Intelligence and the Simulation of Behavior (AISB), 2004 convention: motion. Emotion and cognition, Leeds, 22–26.
- Cappella, J. N., and Greene, J. O. (1982). A Discrepancy-arousal Explanation of Mutual Influence in Expressive Behavior for Adult and Infant-adult Interaction. *Commun. Monogr.* 49 (2), 89–114. doi:10.1080/03637758209376074
- Cappella, J. N. (1991). Mutual Adaptation and Relativity of Measurement. *Studying interpersonal interaction* 1, 103–117. doi:10.1111/j.1468-2885.1991.tb00002.x
- Cappella, J. N. (1981). Mutual Influence in Expressive Behavior: Adult-Adult and Infant-Adult Dyadic Interaction. *Psychol. Bull.* 89 (1), 101–132. doi:10.1037/0033-2909.89.1.101
- Cassell, J., Bickmore, T., Vilhjálmsson, H., and Yan, H. (2000). “More Than Just a Pretty Face: Affordances of Embodiment,” in International Conference of Intelligent Virtual Agents, January 2000 (Springer), 52–59.
- Castellano, G., Pereira, A., Leite, I., Paiva, A., and McOwan, P. W. (2009). “Detecting User Engagement with a Robot Companion Using Task and Social Interaction-Based Features,” in Proceedings of the 2009 international conference on Multimodal interfaces, November 2009 (Cambridge, MA: ACM), 119–126.
- Changchun Liu, C., Conn, K., Sarkar, N., and Stone, W. (2008). Online Affect Detection and Robot Behavior Adaptation for Intervention of Children with Autism. *IEEE Trans. Robot.* 24 (4), 883–896. doi:10.1109/tro.2008.2001362
- Chartrand, T. L., and Bargh, J. A. (1999). The Chameleon Effect: The Perception-Behavior Link and Social Interaction. *J. Personal. Soc. Psychol.* 76 (6), 893–910. doi:10.1037/0022-3514.76.6.893
- Condon, W. S., and Ogston, W. D. (1971). “Speech and Body Motion Synchrony of the Speaker-Hearer,” in *The Perception of Language*. Editors D. L. Horton and J. J. J. (Columbus, Ohio: Charles Merrill), 150–184.
- Coninx, A., Baxter, P., Oleari, E., Bellini, S., Bierman, B., Henkemans, O. B., et al. (2016). Towards Long-Term Social Child-Robot Interaction: Using Multi-Activity Switching to Engage Young Users. *J. Human-Robot Interaction* 5 (1), 32–67.
- Corrigan, L. J., Peters, C., Küster, D., and Castellano, G. (2016). “Engagement Perception and Generation for Social Robots and Virtual Agents,”. *Toward Robotic Socially Believable Behaving Systems*. Editors A. Esposito and L. C. Jain (Springer), 29–51. doi:10.1007/978-3-319-31056-5\_4
- Cuddy, A. J. C., Fiske, S. T., and Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the Bias Map. *Adv. Exp. Soc. Psychol.* 40, 61–149. doi:10.1016/s0065-2601(07)00002-0

- Dermouche, S., and Pelachaud, C. (2019). "Engagement Modeling in Dyadic Interaction," in 2019 International Conference on Multimodal Interaction, October 2019 (Suzhou, Jiangsu: ACM), 440–445.
- Dindia, K. (1988). A Comparison of Several Statistical Tests of Reciprocity of Self-Disclosure. *Commun. Res.* 15 (6), 726–752. doi:10.1177/009365088015006004
- Duchenne, B. (1990). *The Mechanism of Human Facial Expression or an Electro-Physiological Analysis of the Expression of the Emotions*. New York: Cambridge University Press. (Original work published 1862).
- Ekman, P., Friesen, W., and Hager, J. (2002). *Facial Action Coding System (FACS). A Human Face*. Salt Lake City: Research Nexus.
- Fischer-Lokou, J., Martin, A., Guéguen, N., and Lamy, L. (2011). Mimicry and Propagation of Prosocial Behavior in a Natural Setting. *Psychol. Rep.* 108 (2), 599–605. doi:10.2466/07.17.21.pr0.108.2.599-605
- Fiske, S. T., Cuddy, A. J. C., and Glick, P. (2007). Universal Dimensions of Social Cognition: Warmth and Competence. *Trends Cognitive Sciences* 11 (2), 77–83. doi:10.1016/j.tics.2006.11.005
- Gallois, C., Ogay, T., and Giles, H. (2005). "Communication Accommodation Theory: A Look Back and a Look Ahead," in *Theorizing about Intercultural Communication*. Editor W. B. Gudykunst (Thousand Oaks, CA: SAGE), 121–148.
- Giles, H., Coupland, N., and Coupland, J. (1991). "Accommodation Theory: Communication, Context, and Consequence," in *Studies in Emotion and Social Interaction. Contexts of Accommodation: Developments in Applied Sociolinguistics*. Editors H. C. J. Giles and N. Coupland (Cambridge University Press), 1, 1–68. doi:10.1017/cbo9780511663673.001
- Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., Das, M., and Breazeal, C. (2016). "Affective Personalization of a Social Robot Tutor for Children's Second Language Skills," in Thirtieth AAAI Conference on Artificial Intelligence, February 2016 (Phoenix, Arizona: ACM), 3951–3957. doi:10.5555/3016387.3016461
- Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *Am. sociological Rev.* 25 (2), 161–178. doi:10.2307/2092623
- Gueguen, N., Jacob, C., and Martin, A. (2009). Mimicry in Social Interaction: Its Effect on Human Judgment and Behavior. *Eur. J. Soc. Sci.* 8 (2), 253–259.
- Hale, J. L., and Burgoon, J. K. (1984). Models of Reactions to Changes in Nonverbal Immediacy. *J. Nonverbal Behav.* 8 (4), 287–314. doi:10.1007/bf00985984
- Hemminahaus, J., and Kopp, S. (2017). "Towards Adaptive Social Behavior Generation for Assistive Robots Using Reinforcement Learning," in 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Vienna, Austria, 6–9 March 2017 (IEEE), 332–340.
- Huang, L., Morency, L.-P., and Gratch, J. (2010). "Learning Backchannel Prediction Model from Parasocial Consensus Sampling: a Subjective Evaluation," in International Conference on Intelligent Virtual Agents, Philadelphia, PA, USA, September 20–22 (Springer), 159–172. doi:10.1007/978-3-642-15892-6\_17
- Huang, L., Morency, L.-P., and Gratch, J. (2011). "Virtual Rapport 2.0," in International Conference on Intelligent Virtual Agents, Reykjavik, Iceland, September 15–17, 2011 (Springer), 68–79. doi:10.1007/978-3-642-23974-8\_8
- Infante, D. A., Rancer, A. S., and Avtgis, T. A. (2010). *Contemporary Communication Theory*. IA: Kendall Hunt Dubuque.
- Jones, E. E., and Pittman, T. S. (1982). Toward a General Theory of Strategic Self-Presentation. *Psychol. Perspect.* self 1 (1), 231–262.
- Judd, C. M., James-Hawkins, L., Yzerbyt, V., and Kashima, Y. (2005). Fundamental Dimensions of Social Judgment: Understanding the Relations between Judgments of Competence and Warmth. *J. Personal. Soc. Psychol.* 89 (6), 899–913. doi:10.1037/0022-3514.89.6.899
- Katehakis, M. N., and Veinott, A. F., Jr (1987). The Multi-Armed Bandit Problem: Decomposition and Computation. *Mathematics OR* 12 (2), 262–268. doi:10.1287/moor.12.2.262
- Kopp, S., Gesellensetter, L., Krämer, N. C., and Wachsmuth, I. (2005). "A Conversational Agent as Museum Guide - Design and Evaluation of a Real-World Application," in International Conference on Intelligent Virtual Agents, Kos, Greece, September 12–14 (Springer), 329–343. doi:10.1007/11550617\_28
- Lafferty, J. D., McCallum, A., and Pereira, F. C. N. (2001). "Conditional Random fields: Probabilistic Models for Segmenting and Labeling Sequence Data," in Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01, June 2001 (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc), 282–289.
- Lakin, J. L., and Chartrand, T. L. (2003). Using Nonconscious Behavioral Mimicry to Create Affiliation and Rapport. *Psychol. Sci.* 14 (4), 334–339. doi:10.1111/1467-9280.14481
- Levitan, R. (2013). "Entrainment in Spoken Dialogue Systems: Adopting, Predicting and Influencing User Behavior," in Proceedings of the 2013 NAACL HLT Student Research Workshop, Atlanta, Georgia, June 2013 (Atlanta, Georgia: Association for Computational Linguistics), 84–90.
- Lisetti, C., Amini, R., Yasavur, U., and Rishe, N. (2013). I Can Help You Change! an Empathic Virtual Agent Delivers Behavior Change Health Interventions. *ACM Trans. Manage. Inf. Syst.* 4 (4), 1–28. doi:10.1145/2544103
- Lubold, N., Walker, E., and Pon-Barry, H. (2016). "Effects of Voice-Adaptation and Social Dialogue on Perceptions of a Robotic Learning Companion," in 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, New Zealand, 7–10 March 2016 (IEEE), 255–262. doi:10.1109/HRI.2016.7451760
- Maricchiolo, F., Gnisci, A., Bonaiuto, M., and Ficca, G. (2009). Effects of Different Types of Hand Gestures in Persuasive Speech on Receivers' Evaluations. *Lang. Cogn. Process.* 24 (2), 239–266. doi:10.1080/01690960802159929
- Mills, C., Bosch, N., Krasich, K., and D'Mello, S. K. (2019). "Reducing Mind-Wandering during Vicarious Learning from an Intelligent Tutoring System," in International Conference on Artificial Intelligence in Education, Chicago, IL, USA, June 25–29, 2019 (Springer), 296–307. doi:10.1007/978-3-030-23204-7\_25
- Nomura, T., Kanda, T., and Suzuki, T. (2006). Experimental Investigation into Influence of Negative Attitudes toward Robots on Human-Robot Interaction. *AI Soc.* 20 (2), 138–150. doi:10.1007/s00146-005-0012-7
- Paiva, A., Leite, I., Boukricha, H., and Wachsmuth, I. (2017). Empathy in Virtual Agents and Robots. *ACM Trans. Interact. Intell. Syst.* 7 (3), 1–40. doi:10.1145/2912150
- Pecune, F., Cafaro, A., Chollet, M., Philippe, P., and Pelachaud, C. (2014). "Suggestions for Extending SAIBA with the VIB Platform," in Workshop on Architectures and Standards for IVAs, held at the '14th International Conference on Intelligent Virtual Agents (IVA 2014), Boston, USA, August 26 (Boston, USA: Bielefeld eCollections), 16–20. doi:10.2390/biecoll-wasiva2014-03
- Peeters, G., and Czapinski, J. (1990). Positive-negative Asymmetry in Evaluations: The Distinction between Affective and Informational Negativity Effects. *Eur. Rev. Soc. Psychol.* 1 (1), 33–60. doi:10.1080/14792779108401856
- Pelachaud, C. (2009). Modelling Multimodal Expression of Emotion in a Virtual Agent. *Phil. Trans. R. Soc. B* 364 (1535), 3539–3548. doi:10.1098/rstb.2009.0186
- Pennebaker, J. W. (2011). The Secret Life of Pronouns. *New Scientist* 211 (2828), 42–45. doi:10.1016/s0262-4079(11)62167-2
- Raffard, S., Salesse, R. N., Bortolon, C., Bardy, B. G., Henriques, J., Marin, L., et al. (2018). Using Mimicry of Body Movements by a Virtual Agent to Increase Synchronization Behavior and Rapport in Individuals with Schizophrenia. *Sci. Rep.* 8, 1–10. doi:10.1038/s41598-018-35813-6
- Ritschel, H., Baur, T., and André, E. (2017). "Adapting a Robot's Linguistic Style Based on Socially-Aware Reinforcement Learning," in Robot and Human Interactive Communication (RO-MAN), 2017 26th IEEE International Symposium, Lisbon, Portugal, 28 Aug.-1 Sept. 2017 (IEEE), 378–384. doi:10.1109/ROMAN.2017.8172330
- Rizzo, A., Shilling, R., Forbell, E., Scherer, S., Gratch, J., and Morency, L.-P. (2016). "Autonomous Virtual Human Agents for Healthcare Information Support and Clinical Interviewing," in *Artif. intelligence Behav. Ment. Health Care*. Editor D. D. Luxton (San Diego: Academic Press), 53–79. doi:10.1016/b978-0-12-420248-1.00003-9
- Rosenberg, S., Nelson, C., and Vivekananthan, P. S. (1968). A Multidimensional Approach to the Structure of Personality Impressions. *J. Personal. Soc. Psychol.* 9 (4), 283–294. doi:10.1037/h0026086
- Schröder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., ter Maat, M., McKeown, G., Pammi, S., Pantic, M., Pelachaud, C., Schuller, B. W., de Sevin, E., Valstar, M. F., and Wöllmer, M. (2015). "Building Autonomous Sensitive Artificial Listeners," in 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), Xi'an, China, 21–24 Sept. 2015 (IEEE), 456–462. doi:10.1109/ACII.2015.7344610
- Sidner, C. L., Bickmore, T., Nooraie, B., Rich, C., Ring, L., Shayganfar, M., et al. (2018). Creating New Technologies for Companionable Agents to Support Isolated Older Adults. *ACM Trans. Interact. Intell. Syst.* 8 (3), 1–27. doi:10.1145/3213050

- Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., Williams, J., Leuski, A., Narayanan, S., Piepol, D., et al. (2010). "Ada and grace: Toward Realistic and Engaging Virtual Museum Guides," in International Conference on Intelligent Virtual Agents, Philadelphia, PA, USA, September 20-22 (Springer), 286-300. doi:10.1007/978-3-642-15892-6\_30
- Toma, C. L. (2014). Towards Conceptual Convergence: An Examination of Interpersonal Adaptation. *Commun. Q.* 62 (2), 155-178. doi:10.1080/01463373.2014.890116
- van Vugt, H. C., Hoorn, J. F., Konijn, E. A., and de Bie Dimitriadou, A. (2006). Affective Affordances: Improving Interface Character Engagement through Interaction. *Int. J. Human-Computer Stud.* 64 (9), 874-888. doi:10.1016/j.ijhcs.2006.04.008
- van Waterschoot, J., Bruijnes, M., Flokstra, J., Reidsma, D., Davison, D., Theune, M., and Heylen, D. (2018). "Flipper 2.0," in International Conference on Intelligent Virtual Agents, November 2018 (Springer), 43-50. doi:10.1145/3267851.3267882
- Verberne, F. M. F., Ham, J., Ponnada, A., and Midden, C. J. H. (2013). "Trusting Digital Chameleons: The Effect of Mimicry by a Virtual Social Agent on User Trust," in International Conference on Persuasive Technology, Sydney, NSW, Australia, April 3-5 (Sydney, NSW, Australia: ACM), 234-245. doi:10.1007/978-3-642-37157-8\_28
- Wang, C., Biancardi, B., Mancini, M., Cafaro, A., Pelachaud, C., Pun, T., and Chanel, G. (2019). "Impression Detection and Management Using an Embodied Conversational Agent," in International Conference on Human-Computer Interaction, Orlando, FL, USA, 26-31 July (Springer), 392-403. doi:10.1007/978-3-030-49062-1\_18
- Weber, K., Ritschel, H., Aslan, I., Lingensfelser, F., and André, E. (2018). "How to Shape the Humor of a Robot-Social Behavior Adaptation Based on Reinforcement Learning," in Proceedings of the 20th ACM International Conference on Multimodal Interaction, October 2018 (Boulder, Colorado: ACM), 154-162. doi:10.1145/3242969.3242976
- Wiering, M. A. (2005). "Qv (Lambda)-learning: A New On-Policy Reinforcement Learning Algorithm," in Proceedings of the 7th european workshop on reinforcement learning, 17-18.
- Wiggins, J. S. (1979). A Psychological Taxonomy of Trait-Descriptive Terms: The Interpersonal Domain. *J. Personal. Soc. Psychol.* 37 (37), 395-412. doi:10.1037/0022-3514.37.3.395
- Wojciszke, B., and Abele, A. E. (2008). The Primacy of Communion over agency and its Reversals in Evaluations. *Eur. J. Soc. Psychol.* 38 (7), 1139-1147. doi:10.1002/ejsp.549
- Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., and Picard, R. (2009). Affect-aware Tutors: Recognising and Responding to Student Affect. *Int. J. Learn. Tech.* 4 (3-4), 129-164. doi:10.1504/ijlt.2009.028804
- Yzerbyt, V., Provost, V., and Corneille, O. (2005). Not Competent but Warm... Really? Compensatory Stereotypes in the French-speaking World. *Group Process. Intergroup Relations* 8 (3), 291-308. doi:10.1177/1368430205053944
- Zhang, Z., Bickmore, T. W., and Paasche-Orlow, M. K. (2017). Perceived Organizational Affiliation and its Effects on Patient Trust: Role Modeling with Embodied Conversational Agents. *Patient Educ. Couns.* 100 (9), 1730-1737. doi:10.1016/j.pec.2017.03.017
- Zhao, R., Sinha, T., Black, A. W., and Cassell, J. (2016). "Socially-aware Virtual Agents: Automatically Assessing Dyadic Rapport from Temporal Patterns of Behavior," in International Conference on Intelligent Virtual Agents, Los Angeles, CA, USA, September 20-23, 2016 (Springer), 218-233. doi:10.1007/978-3-319-47665-0\_20

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Biancardi, Dermouche and Pelachaud. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.