1 Adaptive Artificial Personalities

Kathrin Janowski, Hannes Ritschel and Elisabeth André

1.1 Motivation

People are known to ascribe human-like social properties to computer systems [Reeves and Nass 1996b]. Users shout at the screen when they are frustrated with the software, while science fiction authors keep envisioning emotional androids and sarcastic A.I. assistants. Various start-up companies work towards bringing artificial beings with endearing personalities into our homes, such as EmoTech's Olly[1] whose adaptive personality is the most emphasized feature in advertising the product.

Socially intelligent assistants, be they disembodied voices or physical robots moving through the shared environment, are expected to become a common sight, especially in the context of care [Janowski et al. 2018]. Equipping such agents with a compelling personality is one important step towards establishing a relationship between human and machine [Breazeal 2004], fostering trust and encouraging long-term use after the novelty effect wears off.

Personality in this context means a person's disposition to respond to certain events in a particular manner [Argyle and Little 1972]. Examples from theater [Laurel 1997] or animated films [op den Akker and Bruijnes 2012] have taught us that a character's reactions to events and how they line up with their goals are what ensures the engagement of human observers. Laurel [1997] argues that giving an interface human-like dispositions helps humans to predict its behavior. These dispositions, in turn, are expressed through and inferred from behavior cues such as interruptions [Rogers and Jones 1975] and gaze behavior [Argyle and Cook 1976].

In order to be accepted, the agent's personality must be tailored to its role as well as the user's preferences for the agent's behavior [Laurel 1997]. Moreover, Aly and Tapus [2016] point out that adapting a robot's personality to the human's profile makes interaction more engaging. However, personality preferences are diverse. Demographic background may influence whether a given phrasing is perceived as convincing or polite [Hammer et al. 2016], while the user's own personality traits may lead them to prefer similar traits in some cases [Bernier and Scassellati 2010, Bickmore and Cassell 2005] and opposite traits in others.

[1] https://www.indiegogo.com/projects/olly-the-first-home-robot-with-personality#/

Furthermore, there are also insights indicating that the task context plays a key role for whether a similar or opposing personality is preferred [Joosse et al. 2013].

But since there is no one true approach for configuring personality just right, enabling a computer system to dynamically adapt to its user is necessary for ensuring effective collaboration between humans and machines. This is why this chapter will cover both configuration before the interaction and adaptation at runtime.

In this chapter, we will consider both virtual agents and social robots. Research in these two fields overlaps greatly, since the personality models and adaptation approaches we will describe can be used for both graphics-based and physically embodied agents. In fact, there are only two key points where the agent types differ. Unlike virtual characters, robots are physically limited in their animation capabilities, for example because their motors can only move at a certain speed or because some postures might make a bipedal robot fall over. The other difference is the physical presence in a shared environment, which may impact the perceived social presence of a robot. However, with recent developments in 3D displays, virtual and augmented reality hardware, we expect this difference to become less pronounced as virtual agents gain similar potential to inhabit the same world as their users.

The chapter is structured as follows. First, section 1.2 will introduce the most important psychological concepts. We will outline commonly used models for personality and interpersonal stance as well as theories about interpersonal compatibility. Next, section 1.3 will describe approaches for implementing adaptive personalities. There, we will summarize ways to express a particular personality through an agent's behavior, and look at different approaches for adapting said personality to the user. After that, section 1.4 will look at common approaches for evaluating user-adaptive systems. Section 1.5 will then give an overview over the historical development in the field of personality adaptation, and section 1.6 will look into current challenges and related future directions. Finally, section 1.7 will summarize and conclude the chapter.

1.2 Psychological Background

For enabling a computer to reason about the personality of the agent and/or the user, the abstract concept needs to be quantified. Psychological literature provides several well-established models for describing personality and the related interpersonal relationships in dimensional terms. This section will present a selection of models that have been successfully applied in human-computer interaction research.

Furthermore, there are several theories regarding the compatibility between and preference for particular personality traits. They build on the aforementioned personality models and will therefore be presented after them.

1.2.1 Personality and Interpersonal Stance

The term *Personality* refers to an individual's behavior patterns that can be observed in a wide range of contexts, or their disposition to respond in a certain way when they find themselves in a particular situation [Argyle and Little 1972]. It manifests in their behavior towards other people, and as this section will explain, links exist between the models used to describe personality and interpersonal stance.

1.2.1.1 Five Factor Model

One of the most widely-used frameworks for describing personality is the so-called *Five Factor Model*, also known as the *Big Five* [McCrae and John 1992, Mehrabian 1996b]. According to this model, personality is defined by the following five dimensions:

- **Openness:** This factor encompasses curiosity, creativity and intellectuality. Open-minded people have a wide range of interests and think in unconventional ways, whereas closed-minded people are generally conservative and unimaginative.
- **Conscientiousness:** This factor is concerned with disciplined and responsible qualities. People who score high in Conscientiousness are generally dutiful and reliable, well-organized and thorough. They are efficient and productive rather than lazy and self-indulgent.
- **Extraversion:** This factor is associated with outgoing and assertive behavior traits. Extraverted people tend to be sociable, talkative, expressive and active, whereas introverts are more reserved and quiet.
- **Agreeableness:** This factor is related to getting along well with others. It encompasses qualities such as being kind, compassionate and forgiving, as well as generous and trusting. In contrast, disagreeable persons tend to criticize others and be cold-hearted or inconsiderate.
- **Neuroticism:** This factor is also known as *Emotional Stability*. It describes a person's tendency to experience negative affect such as distress and anxiety, act impulsively or change moods quickly and frequently. Emotionally stable people, however, are calm and relaxed.

1.2.1.2 PAD Temperament Model

An alternative approach to defining a person's general behavior tendencies is Mehrabian's *Pleasure-Arousal-Dominance Model*, also called the *PAD Temperament Model* [Mehrabian 1996a]. As Mehrabian explained it, the ”temperament” of a person corresponds to their average emotional state across a representative sample of situations, in contrast to the temporary emotional states that change frequently throughout the day. The PAD space is defined by three dimensions which relate to the Big Five factors as follows [Mehrabian 1996b]:

- **Pleasure:** This trait describes a person's tendency to experience positive versus negative emotions. It can be calculated as $0.21 \times Extraversion + 0.59 \times Agreeableness - 0.19 \times Neuroticism$.
- **Arousal:** This trait describes a person's responsiveness to stimuli and the time it takes for them to calm down. According to Mehrabian, it equals $0.15 \times Openness + 0.30 \times Agreeableness + 0.57 \times Neuroticism$.
- **Dominance:** This trait describes how much a person feels in control of their life. Its equals $0.25 \times Openness + 0.17 \times Conscientiousness + 0.60 \times Extraversion - 0.32 \times Agreeableness$.

In computer science, PAD space is also used to model a character's mood or short-term emotions, since these concepts are closely coupled with personality [Gebhard 2007]. Any of them can shape a character's behavior at a given time, so having a common framework for emotions and personality is important for generating consistent system reactions. For more information on emotions see the "Emotion" chapter by Broekens.

1.2.1.3 Interpersonal Circumplex

Attitudes towards other persons are commonly modeled with the *Interpersonal Circumplex* [DeYoung et al. 2013, Horowitz et al. 2006, McCrae and Costa 1989]. It is defined by two axes, *Status* and *Affiliation*.

- **Status:** This axis ranges from *submissive* to *dominant* and is usually displayed as the vertical dimension. It is also known as *Agency* and describes a person's tendency to act according to their own will.
- **Affiliation:** This axis ranges from *cold* to *warm* and is placed horizontally. Also known as *Communion*, it describes a person's social closeness to other people.

These two dimensions have also been shown to be related to the personality traits Extraversion and Agreeableness [DeYoung et al. 2013, McCrae and Costa 1989]. As explained in section 1.2.1.1, high Extraversion implies sociability and therefore closeness to people, but also assertive and therefore dominant behavior tendencies. In a similar manner, Agreeableness represents social compliance which is a combination of warm-hearted and submissive behavior. According to literature [DeYoung et al. 2013, McCrae and Costa 1989], these two personality traits form an alternate pair of axes which is rotated about 30-45° relative to status and Affiliation, as shown in figure 1.1.

1.2.1.4 Politeness Theory

Another important concept for interpersonal behavior is politeness. In their *Politeness Theory*, Brown and Levinson [Brown and Levinson 1987] assume that every human has two basic

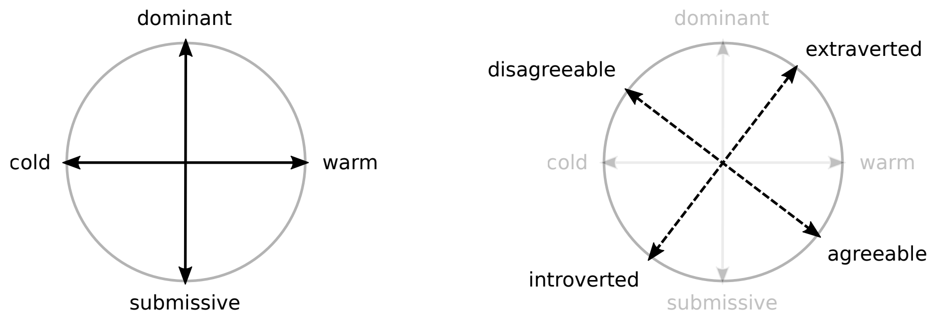

Figure 1.1 The two pairs of dimensions which define the Interpersonal Circumplex. *Left/solid:* Status and Affiliation. *Right/dashed:* Extraversion and Agreeableness.

wants concerning their public identity, also called their *face*. Those wants, in turn, appear related to the Interpersonal Circumplex [Oakman et al. 2003].

- **Negative Face Want:** People desire to be autonomous in their actions. This resembles the Status dimension of the Interpersonal Circumplex which represents a person's tendency to act autonomously.
- **Positive Face Want:** This concerns the desire to have other people's approval and know that they share one's own goals. Like the Affiliation dimension, it implies group membership and a social bond with others.

Different phrasings can be categorized according to the *face threats* they present or avoid. Accordingly, those that minimize threats to the hearer's positive face are said to use *positive politeness*, and those that minimize negative face threats use *negative politeness*. For example, Johnson et al. examined eight different phrasings in English and German, and found that the categories listed in table 1.1 are interpreted as shown in figure 1.2 [Johnson et al. 2005].

1.2.2 Theories about Interpersonal Compatibility

There are two major theories concerning the compatibility of individuals based on their personality. While the *similarity attraction* theory suggests that people would be most compatible with similar personalities ("birds of a feather flock together"), the *complementarity* theory suggests that people are more compatible with dissimilar personalities ("opposites attract"). Additionally, there are approaches to reconcile both ideas by considering the underlying goals which shape either person's behavior and appraisal of events.

1.2.2.1 Similarity Attraction Theory

Similarity attraction is most often researched with regards to attitudes, but comparable effects have been found for personality traits as well [Montoya and Horton 2013].

phrasing	**example**
direct command	”Drink some tea.”
indirect suggestion	”The system is asking you to drink some tea.”
request	”I would like you to drink some tea.”
system’s goal	”I would drink some tea.”
shared goal	”We should drink some tea.”
question	”How about drinking some tea?”
suggestion of user’s goal	”You would probably like to some tea.”
socratic hint	”Did you think about drinking some tea?”

Table 1.1 Categories of phrasings with different degrees of positive and negative politeness, according to Johnson et al. [Johnson et al. 2005].

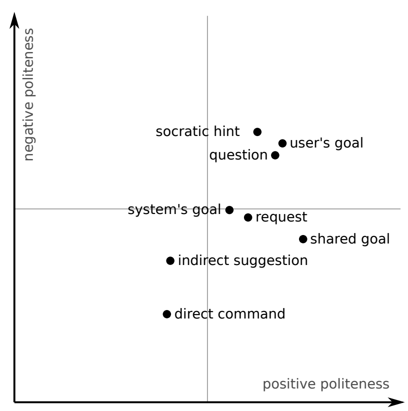

Figure 1.2 The perceived positive and negative politeness of different phrasings, according to Johnson et al. [Johnson et al. 2005].

Moon examined the effect of dominance on the persuasiveness of computer-generated messages, depending on the dominance of the human interacting with it [Moon 2002]. Dominant messages were phrased as assertions and direct commands while expressing a high confidence (given in percent) in the computer’s claim. Submissive messages, in contrast, used questions and suggestions in combination with low confidence levels. The results showed that dominant users were more likely to change their ranking of different cars when the computer used a dominant message style to contradict them. They also rated the information quality higher. When asked to judge the computer’s level of expertise, all participants interacting with one matching their own dominance rated the computer more favorably. A follow-up experiment showed that participants rated music samples, cartoons and health tips more favorably when the computer's presentation style matched their own dominance level. An earlier study by Moon and Nass [1996], also using verbal cues and confidence levels to manipulate dominance, showed that people preferred computers whose dominance level became more like theirs over time, rather than those staying the same or developing in the opposite direction.

Andrist et al. [2015] varied a robot's gaze behavior to indicate either an introverted or extraverted personality. They then had people collaborate with the robot on a puzzle task, leaving the subjects to decide when they wanted to stop. Their results showed that people without intrinsic motivation to solve the puzzle spent significantly more time collaborating with a robot whose Extraversion level was similar to their own, and chose to solve more puzzle tasks together.

1.2.2.2 Complementarity Theory

Behaviors which are elicited by one reflecting a particular combination of status and affiliation (see section 1.2.1.3) are called complementary [Estroff and Nowicki Jr. 1992, Markey et al. 2003].

Markey et al. observed the behavior during the encounter, collaborative and competitive interaction of unacquainted dyads [Markey et al. 2003]. They confirmed that the model which best explained their observations was the one where behaviors elicited reactions that were similar in affiliation, but opposite in status.

Estroff et al. examined whether complementary pairs of people were more effective in performing puzzle or word finding tasks than anti-complementary pairs [Estroff and Nowicki Jr. 1992]. They used the definition that complementary people would be similar with regards to the affiliation dimension, but opposite with regards to status. For anti-complementary pairs, the dimensions were switched. Their study results showed that complementary dyads completed a significantly greater part of the puzzle. For the word finding task, the tendency was the same but not significant.

Similar patterns could be confirmed for the Big Five personality trait of Extraversion. As section 1.2.1.3 explained, high Extraversion is closely related to high dominance. Liew and Tan [2016] found that learners reported more positive emotions and learning motivation when interacting with a virtual tutor whose Extraversion was the opposite of their own. Extraverted learners also found the introverted agent more likeable and trustworthy.

1.2.2.3 Interpersonal Goals

Since neither similarity attraction nor complementarity are universally applicable, other researchers have focused on the underlying interaction goals and personal motivations.

For instance, Tett and Murphy [2002] proposed that "people prefer co-workers who let them be themselves", which would explain both the similarity attraction and the complementarity theory. According to them, expressing one's personality traits is a fundamental human need (which is also supported by the face wants of the politeness theory, see section 1.2.1.4). Consequently, similarity is preferred for agreeableness or affiliation because it enables people to express mutual closeness, whereas complementarity is preferred for dominance when one person wants to lead and the other wants to follow.

Horowitz et al. [2006] explained that a person's goal can be related to status or affiliation, and that frustration arises when the other party misinterprets that goal. One example they give is that a person might talk about a problem seeking either guidance (status-based motive) or comfort (affiliation-based motive). The interaction breaks down if they receive something different from what they expected.

Reisz et al. [2013] examined how a person's goals relate to their personality traits. They found that people formulate goals which lead to positive affect or compensate for negative affect, especially if those goals require great or continued effort. For example, open-minded people named the goal of learning a new skill, whereas introverted people named the goal of making new friends. Goals like "use time more effectively" were given by people low in conscientiousness, whereas "reduce stress" was listed by those scoring high on that trait and on neuroticism. This explanation implies that people are compatible with others who reinforce their positive traits (such as two friends high in agreeableness) or compensate for a deficit in one trait (such as introverted people being friends with less introverted persons).

1.3 Implementing Adaptive Systems

There are two main tasks when implementing personality-adaptive systems: One is the development of an adaptation mechanism, and one is the mapping between personality traits and observable surface cues. We will start with the latter because adaptation mechanisms require an understanding of the parameters that will be modified over time.

1.3.1 Expressing Personality with SIAs

Personality can be expressed with multiple communication channels due to the flexibility of today's virtual agents and social robots. These expressions are commonly adapted from psychological findings about human communication, such as those summarized by Knapp et al. [2014].

First of all, the *linguistic content* of utterances can be used to express personality. For example, Moon and Nass [1996] use scripted language-based cues (strong language, assertions and commands to express high confidence vs. weaker language, questions and suggestions to express low confidence) to make the computer appear either dominant or submissive. The PERSONAGE natural language generator by Mairesse and Walker [2010] realizes linguistic variation according to the Big Five personality dimensions in the context of restaurant comparisons, which is used by Aly and Tapus [2016] with a social robot. Ritschel and André [2017], Ritschel et al. [2017] use an adaptive, generative approach for a story telling robot with varying degree of extraversion. Harrison et al. [2019], Hu et al. [2018] present neural generation methods with variations in personality (Big Five) and improved stylistic control in the context of task-oriented dialogs. *Paralinguistic behavior* is of central importance to present the linguistic content in a convincing manner. For example, synthesized voices indicate extraversion with faster speech rate, higher pitch and volume, and wider range of tone; in contrast, slower speech rate, lower pitch and volume, and smaller range of tone indicates introversion [Reeves and Nass 1996a].

Kim et al. [2008] and Hu et al. [2015] manipulate a robot's *gestures* to express extraversion and introversion. Their agents' gestures vary in amplitude, speed, frequency, and more. Similarly, Isbister and Nass [2000] express this dimension with a virtual character with *posture*. This includes wide spread limbs, wider movement and movement towards the observer for extraversion, and limbs closer to the body and less freely gestures for introversion. Generative approaches [Hartmann et al. 2005] are important for creating this type of behavior flexibly, without the need of hardcoding and scripting every detail of the interaction. Given enough personality-related training data from audio corpora, future applications may also make use of deep learning based speech-driven gesture generation [Kucherenko et al. 2020]. In a similar vein, *gaze behavior* can provide clues about an agent's traits and attitudes. For instance, Bee et al. [2009] showed that the orientation of a virtual character's head and eyes influenced the dominance level expressed by the agent. Later, Arellano et al. [2011] showed that agents turning their head upwards were perceived as more extraverted and less agreeable than those lowering their gaze.

Due to the successful personalization [Ritschel et al. 2019a] of artificial agents' nonverbal sounds [Bethel and Murphy 2008] with regard to expression of emotion and intentions, generating such sounds during runtime [de Gorostiza Luengo et al. 2017] is of high interest to shape socially interactive agents' personality beyond the traditional verbal and nonverbal modalities.

Studies with embodied conversational agents have shown that humans also interpret their *turn-taking patterns* in terms of personality and interpersonal attitudes. In general, starting to speak later and yielding to interruptions sooner are signs of an introverted and submissive character, whereas agents which interrupt the other party and continue to talk over them appear extraverted and dominant. These relationships between speech timing and personality were confirmed by ter Maat et al. [2011], by Cafaro et al. [2016], as well as by Janowski and André [2019]. Additionally, Gebhard et al. [2019] showed that the same is true for interactive human-agent conversation.

1.3.2 Adaptation Approaches

Figure 1.3 gives an overview of adaptation and expression of personality in the context of SIAs. In general, one can distinguish between *adaptive* and *adaptable* social agents [Schneider and Kummert 2020]. The former autonomously attempts to adapt to the user, the latter allows the user to actively change the agent's behaviors. In the context of personality, the

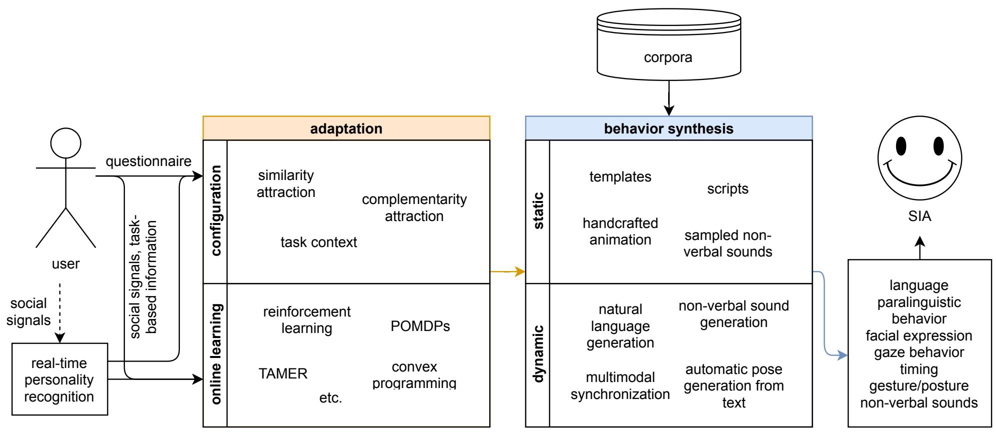

Figure 1.3 Adaptation and expression of SIA personality.

term *adaptation* is used for manipulating an agent's behaviors in order to express a personality which is best for the individual user. This is realized by either configuring the robot's personality profile according to the similarity or complementarity attraction principle (see section 1.2.2), or by tweaking parameters during runtime. Typically, human input is provided either as data upfront to the interaction or online during the interaction. The former is often realized e.g. by filling out a questionnaire or self-report to assess the user's own personality profile. The latter includes approaches for realtime processing of the user's social signals or task-based information, such as user performance in a goal-oriented interaction. For example, paralanguage and nonverbal behavior are used for automatic personality recognition [Vinciarelli and Mohammadi 2014]. Social signal processing uses speech, gestures, pose, facial expression and more to fulfill this task based on human input. This is done to either estimate the human's personality profile or to get other information about the user. For example, Carbonneau et al. [2020] use feature learning and spectrogram analysis for estimating personality via speech and Salam et al. [2017] use a fully automatic system to predict the human Big Five personality traits based on nonverbal behavioural cues.

Social signals can also be used to estimate human engagement. Engagement [Oertel et al. 2020] is of central importance for designing systems, which are able to adapt to their users' characteristics [Salam et al. 2017]. For example, Mancini et al. [2019] use engagement detection to form and maintain a virtual agent's impression of warmth and competence. Since the adaptation of the robot's personality profile can make interaction more engaging (see section 1.2.2) engagement is also investigated in the context of human and robot personality, e.g. in Celiktutan et al. [2017]. In combination with adaptation, Ritschel and André [2017], Ritschel et al. [2017] use it to optimize a social robot's extraversion based on the user's reactions.

1.3.2.1 Decision-theoretic Reasoning

The situations which an agent encounters are not always deterministic. Users might respond positively to a certain personality expression most of the time, but feel irritated by it when they are in a certain mood or busy with a particular task. Therefore, decision-theoretic approaches consider probabilistic outcomes that are linked to a person's internal affective or cognitive states. Since the latter are not directly observable, they need to be inferred from surface behavior cues, for example using a Bayesian Network [Ball and Breese 2000, Conati 2013].

Ball and Breese [2000] used such a Bayesian network to model the causal relationship between different behavior features, the current emotion (corresponding to the first two dimensions of the PAD model) and the interpersonal stance. This model was then applied both for inferring the user's current affective state and for modifying the agent's behavior. The considered modalities included, among others, choice of words, acoustic features such as pitch and volume, timing information and gestures. Furthermore, the authors pointed out that such a model can be easily extended by creating a new chance node for the added modality and setting the conditional probabilities for observing certain behavior depending on the emotional state and interpersonal stance.

Conati [2013] described an approach for implementing an affect-sensitive virtual butler that would adapt the timing of its actions to the user's state. In this case, a *Dynamic Decision Network* (a Bayesian Network augmented with temporal information and decision nodes) serves for inferring the user's current emotion not only from their behavior, but also from background knowledge about their personality-related goals. This enables the system to distinguish between emotions that are similar in valence but are caused by different events and therefore require different responses. For example, if the user felt shame over making a mistake, the virtual butler would need to bolster the former's confidence, whereas if the user felt reproach because of the system's mistake, the agent would need to apologize. In such cases, a socially interactive agent may need to deviate from its default personality, showing more respectively less assertive behavior to accommodate the current interaction context.

One major ingredient of decision-theoretic systems are the costs and benefits associated with different outcomes, which define each outcome's utility. The so-called expected utility is the sum of said utilities, weighted by the probabilities that the system action in question will lead to the respective outcomes. This makes it possible to predict which of the available actions will be the most beneficial in the face of uncertain situations.

For instance, Bohus and Horvitz [2011] had human judges rate the gravity of different turn-taking errors made by a virtual agent in order to create a cost function based on the time that the system chose to wait in different situations. After calculating the systems's confidence that a user had yielded the turn to the virtual agent, this cost function could be used to decide about the ideal waiting time before the agent started speaking. In a similar manner, the virtual butler

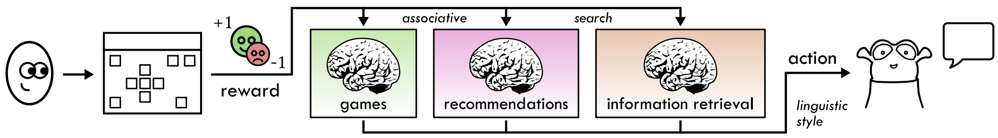

Figure 1.4 Contextual bandit/associative search (i.e., several multi-armed bandit problems) are used to adapt a social robot's linguistic style in Ritschel et al. [2019d].

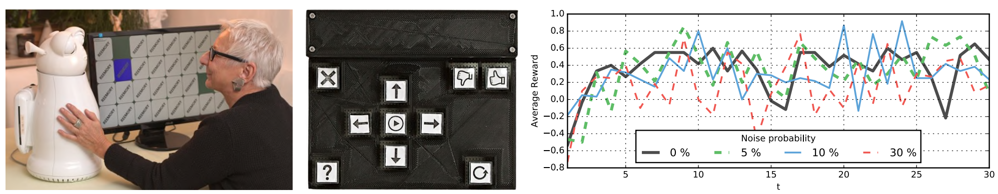

Figure 1.5 Left: a study participant interacts with an adaptive social robot in Ritschel et al. [2019d]. Middle: the hardware control panel is used for interaction and to give a reward signal for reinforcement learning (adapted from Ritschel et al. [2019d]). Right: simulation results of a robot with adaptive extraversion in Ritschel et al. [2017]. Reeti robot by Robopec.

proposed by Conati [2013] would be able to calculate the expected impact that its timing would have on a user's mood.

1.3.2.2 Reinforcement Learning

In recent years, Reinforcement Learning [Sutton and Barto 2018] is becoming increasingly popular, also in the context of behavior adaptation for social agents. It is often used for personalization [den Hengst et al. 2020]: exploring and identifying the best agent's behavior for the individual human interaction partner.

In general, a reinforcement learning agent explores different *actions* in different *states* iteratively. For every action execution it receives a numeric *reward*. This information gives it an indication of whether the action was expedient or not. Since reinforcement learning solves control problems the agent needs to balance *exploitation* (i.e. selecting the "greedy" action, which is the most profitable one according to the agent's experience) and *exploration* (i.e., selection of suboptimal actions) when deciding which action to take for the next learning step. In human-agent interaction, feedback for calculating the reward is provided implicitly or explicitly based on different sources of information, including task-related information and human social or biosignals. Explicit feedback includes e.g. ratings via haptic button presses [Ritschel et al. 2019d] (see Figure 1.5), graphical user interfaces [Ferreira and Lefèvre 2015], tactile [Barraquand and Crowley 2008, Wada and Shibata 2006] or prosodic [Kim and

Scassellati 2007] input. Implicit feedback is not provided consciously by the user but either deducted from task-related information or subconscious human social or biosignals. Task-related data is of central importance to estimate the user's performance in goal-oriented tasks, exercises or games, such as in Ritschel et al. [2018, 2019b, 2020b]. However, it cannot capture human aspects, such as the user's behavior, personality, or mood. Thus, interaction distance, gaze and smile [Fournier et al. 2017, Gordon et al. 2016, Hemminghaus and Kopp 2017, Leite et al. 2011], motion speed, timing [Mitsunaga et al. 2008], gesture and posture [Najar et al. 2016, Ritschel et al. 2017], and laughter [Hayashi et al. 2008, Katevas et al. 2015, Knight 2011, Ritschel et al. 2020a, Weber et al. 2018] are used in various contexts as feedback for social agents. Physiological feedback includes ECG [Liu et al. 2008] or EEG [Tsiakas et al. 2018] data. These signals are often aggregated and combined in order to build a user model and calculate reward, e.g. based on the human's estimated affect/emotions [Broekens and Chetouani 2019, Gordon et al. 2016, Leite et al. 2011], engagement [Mancini et al. 2019, Ritschel 2018, Ritschel et al. 2017], curiosity [Fournier et al. 2017], amusement [Ritschel and André 2018, Ritschel et al. 2020a, Weber et al. 2018] and more.

A range of reinforcement learning algorithms is used for social agent adaptation. The simplest form are multi-armed bandit problems [Leite et al. 2011], duelling bandits [Schneider and Kummert 2020] or contextual bandits/associative search [Ritschel et al. 2019d] (see Figure 1.4). Popular full reinforcement learning algorithms include Q-Learning [Barraquand and Crowley 2008, Broekens and Chetouani 2019, Gordon et al. 2016, Hemminghaus and Kopp 2017, Liu et al. 2008, Ritschel et al. 2017, Tsiakas et al. 2018] and policy gradient reinforcement learning [Mitsunaga et al. 2008, Tapus et al. 2008]. Pietquin [2013] furthermore outlines the use of inverse reinforcement learning for interactive systems, which allows to learn the reward function by observing human experts. The author also emphasizes an important challenge: the phenomenon of coadaptation, which occurs when an agent adapts to the user and the human subsequently adapts to the adapted system.

With regard to personality, Iida et al. [1998] use reinforcement learning for learning personality-related cues for a robot. The human moves through different spatial areas in front of the robot, while each movement is followed by a robot's action. Each of these actions is rewarded positively or negatively. Tapus et al. [2008] use reinforcement learning to optimize the robot's personality during exercises in the context of post-stroke rehabilitation therapy. The authors adjust interaction distance, speed, and vocal content of the therapist robot in order to improve the user's task performance. Introversion is expressed with gentle, supportive language, low pitch and volume, and extraversion uses challenging language, high pitch and volume. The reward signal is based on the number of exercises performed in a given period of time. Ritschel et al. [2017] use reinforcement learning to optimize a social robot's linguistic style in terms of extraversion based on the user's engagement. While the natural language generation approach is inspired by PERSONAGE [Mairesse and Walker 2010] human engagement is estimated based on social signals. An increase/decrease of engagement

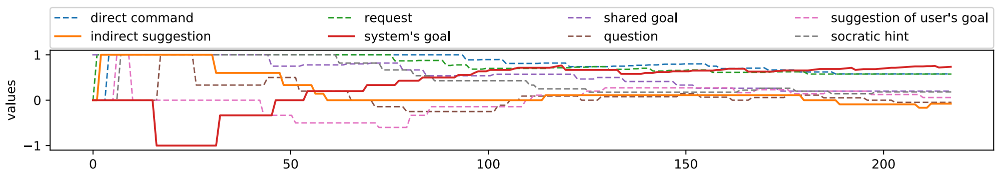

Figure 1.6 Estimated values of a robot's learned politeness strategies over time based on the feedback received from a female test person. Solid lines represent the resulting best (red) and worst (orange) strategy. Changes become smaller over time (modelled as a stationary problem).

over time results in a positive/negative reward for the learning agent. As Martins et al. [2018] point out there is a lack of experiments with adaptation based on a deeper, psychological understanding of users. The authors expect to achieve more user satisfaction and acceptance of social robots when psychological measures, such as user personality, are used for adaptation.

One important aspect in reinforcement learning is the distinction of stationary and non-stationary problems. For example, Figure 1.6 illustrates the values learned by a multi-armed bandit over time based on the feedback it received from the human user. Over time, changes become smaller and smaller (stationary problem), estimating the final values more precisely. As soon as human preferences change over time, it is necessary to use algorithms for non-stationary problems. Typically, a constant, small learning rate is used, which controls to which degree new feedback affects the learned policy to date, i.e., how quickly the agent can react and adjust its knowledge to the human's new desires.

1.3.2.3 Neural Learning of Personality-Based Behavior Styles

While previous approaches generated behavior variants by applying procedural manipulations, more recent approaches employ a data-driven generation paradigm where neural networks are trained on large corpora to automatically synthesize stylistic variants for verbal and nonverbal behaviors.

Nguyen et al. [2018] trained a chatbot on TV show transcripts and movie dialogues to learn how to map user utterances on chatbot responses. To this end, they made use of an encoder-decoder-setup with an attentional mechanism that consisted of an encoder to process user utterances and a decoder to produce the chatbot answer. An evaluation of the chatbot revealed that the approach was able to learn certain aspects of linguistic style that may be attributed to personality.

The approach by Nguyen et al. [2018] focused on the achievement of stylistic goals without considering the semantic content of utterances. Such an approach is, however, not suitable for task-based dialogue. Oraby et al. [2018] investigated how to train a neural model for task-based dialogue that ensures not only stylistic variation, but also semantic fidelity. To acquire a sufficient amount of training material consisting of semantic representations of dialogue acts and matching language output representing different styles, they made use of the PERSONAGE natural language generator by Mairesse and Walker [2010] (see Section 1.3.1). Experiments with various encoder-decoder setups showed the benefit of neural architectures for controlling stylistic variants in task-based dialogue.

While the approaches above aimed to realize explicitly given stylistic goals, Hoegen et al. [2019] made use of neural architectures to implement a chatbot that matches the participant's conversational style on previously defined variables, such as pronoun use and speech rate. Conversational style matching can be regarded as a kind of social adaptation mechanism where conversational partners aligns their conversational behaviors to each other [Schiller et al. 2019]. While there is no unique mapping between conversational style variables and personality traits, there are obvious connections. For example, length of utterances and loudness of speech may be an indicator of extraversion.

Neural network approaches have also been exploited for the generation of stylistic variants for character animations. A particular style of motion modulates how a character is perceived and which emotions, mood and personality are ascribed to it. For example, Smith et al. [2019] trained three separate networks for pose, timing and foot to enable real-time style transfer. Style transfer helps reduce the amount of data including combinations of heterogeneous actions and motion variants required to achieve character animations of sufficient quality that enhance the character's expressivity. Even though the work by Smith et al. [2019] focused on a broader range of affective and non-affective components, the approach bears great promise for generating motion style variants that portray a character's personality.

1.4 Evaluating Adaptive Systems

Evaluating an adaptive system is not trivial because it needs to be observed over an extended period of time. Furthermore, the frame of reference is subject to change, making it difficult to compare results between, for example, different iterations of a prototype. One way to approach this issue is to evaluate components separately, for example testing the learning mechanism with simulated users while conducting a perception study to confirm that the generated behaviors convey the intended personality. This section summarizes the most important evaluation techniques.

1.4.1 Simulation

When it comes to online adaptation mechanisms, simulations are used to investigate whether the implemented algorithms work in theory. Since studies with real users are complex, time consuming and expensive, simulated users are one option to check for algorithmic issues before testing the system in real human-agent experiments (see Section 1.6.1). This is a first important step which should be done in advance since simulations are cheap, repeatable and easily allow for changes and parameter tweaking. For example, Jain et al. [2018] propose a user simulator architecture for dialogue managers, tailored to socially-aware conversational agents, which generates both task and social behaviors.

Typically, a simulated user model or fake user is used to mimic reactions to an agent's actions, such as in case of reinforcement learning (see Section 1.3.2.2). For example, the right subfigure in figure 1.5 illustrates simulation results from Ritschel et al. [2017], where a social robots adapts its extraversion based on the human's engagement. Depending on what is adapted and which human data serves as input for this process, simulations require a strong level of abstraction and typically cannot emulate real human reactions adequately due to the high complexity. One approach to overcome this issue can be a combination with Wizard-of-Oz Studies (see Section 1.4.3), which are used e.g. in spoken dialogue systems [Rieser and Lemon 2011]. Thus, simulations can be used for verifying an adaptation approach's theoretical functionality and for parameter tuning.

1.4.2 Perception Studies

To evaluate personality-based behavior variations, many researchers use perception studies in which they present study participants with videos of pre-recorded interactions between social agents. Examples can be found in the works of Cafaro et al. [2016], ter Maat et al. [2011] Ravenet et al. [2015] or Janowski and André [2019]. One advantage of this approach is that all study participants observe and judge identical situations. It removes the risk of technical problems during the presentation of the agents' behavior, and avoids interference caused by non-deterministic behavior processes (such as random noise in the gaze direction) which are often implemented to make the agents' actions appear more natural.

Additionally, using video stimuli allows researchers to reach more participants from more diverse backgrounds because it removes the need for people to come to the laboratory. Instead, perception studies can be conducted over the internet, using online surveys or crowdsourcing platforms such as Amazon Mechanical Turk.

However, putting the subjects in the role of passive observers can have both positive and negative effects. On the plus side, it allows the users to focus on the agent's behavior rather than the interaction task. On the minus side, this detachment from the task may lead to different evaluations than those from actually using the system themselves. Berry et al. [2009] found that subjects accepted a scheduling assistant's sub-optimal suggestions under laboratory conditions because the scheduled events were purely fictional and people were aware that they would not actually need to attend them. The same effect can be expected when users know they do not have to interact with the agent whose behavior they are rating.

1.4.3 Wizard-of-Oz Studies

Robert Jr. et al. [2020] found that the majority of human-robot interaction studies use Wizard-of-Oz setups or hybrid systems where only part of the robot's behavior is truly autonomous.

The most compelling reason for this is that it allows for rapid prototyping of behaviors, for testing the effect of behavior variations without the need to implement them.

It also mitigates problems with input processing. Having a human observe the user's voice or gesture commands makes it possible to simulate near-perfect recognition of not only the surface behavior, but also the semantics behind it. In contrast, automatic recognition often requires enormous amounts of data and computation resources, as evidenced by the fact that state-of-the art voice interfaces rely on cloud-based services to process the audio input.

However, Wizard-of-Oz studies tend to show an over-idealized version of the system, and findings from such studies can be difficult to transfer to autonomous implementations. Furthermore, special care needs to be taken to ensure that the remote operator (the "wizard") behaves in a consistent manner towards all participants within a given condition. Unlike an autonomous application, human operators are subject to fatigue and distractions. Methods for supporting these wizards in their task, for example by automating part of the robot's behavior, consequently form a research topic of their own.

1.4.4 Autonomous Interactive Systems

Ideally, an agent's behavior is evaluated during autonomous interaction with real users, to prove that the approach is feasible in reality and that the implementation works as expected. This, of course, requires a fully-functional prototype, so this type of evaluation is mainly used near the end of a project.

Furthermore, the users, to whom the system needs to adapt, may behave very differently in their regular environment, as observed by Berry et al. [2009] in the article cited above. Because the users accepted sub-optimal recommendations under laboratory conditions, the meeting scheduling agent was improperly trained. The authors therefore suggest to deploy working implementations of adaptive systems to end-users in order to have them evaluated under real-life conditions. However, they also point out that this requires the basic application to have good usability to begin with, so that people will use it long enough to actually observe the adaptation.

Good usability depends on a number of factors. Among other things, the application needs to have reliable input processing (such as accurate speech recognition in a noisy environment), fallback strategies for handling unexpected user input and an intuitive dialogue structure which guides the user towards their options. Unfortunately, shortcomings in that regard are often only found when the system is confronted with naive end-users. One way to deal with this issue is to incrementally improve the application and re-evaluate each new version, but this takes a considerable amount of time and may alienate users in the early stages.

1.5 History / Overview

After more than a quarter of a century of research on virtual agents, it is common belief that computer characters need to be realized as individuals with a distinct personality. However,

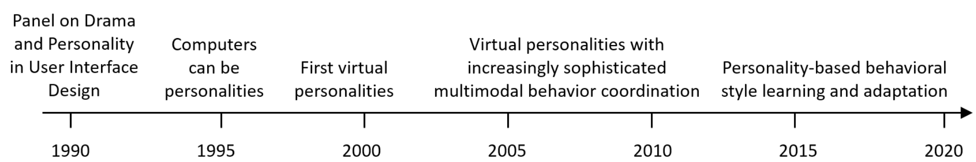

Figure 1.7 Timeline of Research on Virtual Personalities

three decades ago, the idea of crafting a personality for computer systems was considered anything but obvious. Indeed, a panel on drama and personality in user interface design [Mountford et al. 1989] at the 1989 SIGCHI Conference on Human Factors in Computing Systems (see Fig. 1.7) introduced the topic with the words: ”Of all the things that come to mind when one thinks of computers and user interfaces, drama and personality are among the last.” The motivation to discuss of whether and how to give computers a personality was driven by the promise to create new experiential user interfaces.

In 1995, Nass et al. [1995] posed the provocative question: ”Can computer personalities be human personalities?” At that time, first attempts have been made to convey verbal and nonverbal cues by computer agents [Cassell et al. 1994]. However, it was not clear whether people would interpret behavioral cues provided by a computer agent in a similar manner as behavioral cues provided by a human. To shed light on this question, Nass and colleagues conducted a first study with verbal stimuli. The study revealed that simple cues, such as strong versus weak language, already enable a human to attribute a particular personality to a computer agent. They concluded that a minimal set of cues suffices to provide a computer with a personality that increases user liking.

In 1997, the first collective volume on the creation of virtual personalities [Petta and Trappl 1997] appeared. As the early SIGCHI panel it made allusions to drama and theatre and presented a variety of approaches that aimed to create synthetic actors with a personality. A representative example is the work by Hayes-Roth et al. [1997] who implemented several scenarios following the metaphor of a virtual theater. Their characters were not directly associated with a specific personality. Instead, they were assigned a role and had to express a personality that was in agreement with this role.

While Nass and colleagues relied on handcrafted stimuli in their personality experiments, researchers in the area of computer animation and natural language generation started to investigate how to automatically create behaviors for virtual agents that portray a particular personality. Parameters related to personality were used to control many aspects of a virtual character’s multimodal behaviors including movement quality [Badler et al. 1997], linguistic style and prosody [Walker et al. 1997], or dialogue policy [André et al. 2000].

While earlier approaches on virtual agents were inspired by work on drama and theatre, research around the start of the millennium was marked by an increased interest in explicitly modeling theories from the psychological sciences, such as the Big Five [McCrae and John 1992] for personality and the Ortony Clore Collins (OCC) model [Ortony et al. 1988] for emotions. Ball and Breese [2000] treated personality and emotion as unobservable variables in a Bayesian Network and defined model dependencies between these unobservable variables and observable ones, such as linguistic style and facial expressions (see Section 1.3.2.1). André et al. [1999] presented several projects with virtual agents in which personality and emotions were used as filters to constrain the decision process when selecting and instantiating the agents' behaviours. Rist et al. [2003] developed a platform for the realization of multiple conversational settings with virtual agents that portrayed through their dialogue behaviors a particular personality specified by the human user beforehand. These early approaches enabled a flexible generation of multimodal behaviors based on given personality profiles. However, it was not possible to adjust the behaviors of the virtual agents based on the interaction with the user even though first ideas on the evolvement of a virtual character's personality were discussed in [Rist et al. 2004].

Even though first tools to temporally synchronize behaviors for virtual agents, such as PPP (Personalized Plan-Based Presenter, [André et al. 1998]) or BEAT (Behavior Expression Animation Toolkit, [Cassell et al. 2004]), were in place, consistency and timing of behaviors was still considered a major challenge at the beginning of the millennium [Gratch et al. 2002]. Around the same time, a study by Isbister and Nass [2000] emphasized the importance of orchestrating multiple cues, such as postures and language, to portray personality in a consistent manner. Back then usually just one character was used that had to fit the preferences of all users. For this reason, Isbister and Nass [2000] did not rule out the possibility that a character with a partially matching personality might be better than a character with a mismatching personality. Their study showed, however, that users have a strong preference for characters that show a consistent behavior independent of their personality profile. Contrary to previous research, the similarity-attraction hypothesis (see Section 1.2.2) could not be confirmed.

The first decade of the new millennium was characterized by a larger variety of carefully coordinated behaviors that improved the expressiveness of the virtual characters [André and Pelachaud 2010]. This development was fostered by a variety of scheduling approaches, such as SmartBody [Thiébaux et al. 2008] or MARC [Courgeon and Clavel 2013], that automatically assembled synchronized animations and speech based on behavior descriptions in BML (Behavior Markup Language, [Vilhjálmsson et al. 2007]). The more sophisticated behaviors of virtual agents led to studies that investigated the impact of multiple modalities, such as language and gesture [Neff et al. 2010] or language and gaze [Bee et al. 2010], on the perception of personality. Cafaro and colleagues [Cafaro et al. 2012] showed that people already form a first impression of an agent's personality and interpersonal attitude at first encounters based on nonverbal immediacy cues, such smile, gaze and proximity.

The first fifteen years of the new millennium were marked by approaches that created personality profiles for virtual agents beforehand, for example, to conducted perception studies as described above. For the last few years, a trend can be observed, however, to dynamically adjust a virtual agent's personality to a user's (potentially changing) preferences (see Section 1.3.2.2). This work is motivated by the observation that the similarity-attraction hypothesis does not necessarily hold and it is thus hard to predict a user's preferences regarding the personality of a virtual agent. Furthermore, there is a trend towards neural behavior generation approaches that enable the synthesis of a large variety of behavior styles for multiple modalities, such as movement and language (see Section 1.3.2.3). The two trends come with the promise to achieve the next level of human-likeness in the area of virtual agents.

1.6 Current Challenges

This section summarizes current research directions and unanswered questions. In particular, most applications of personality adaptation have yet to leave the laboratory and be tested with real users in their regular environment. The best way to do so would be by deploying fully-functional systems to end users, but in order to develop those, several problems need to be solved first.

1.6.1 In-situ studies

Socially interactive agents are getting closer to being integrated in our everyday life and domestic environments, thanks to many research insights and technological advances in the last decade. With human preferences being diverse, an investigation of (adaptive) social agents in the field of application, i.e. "in the wild", is important. However, the evaluation of social agents poses a number of challenges, in particular when targeting elderly people, domestic environments, and adaptation (see Figure 1.5). Users, who did not use computing technologies throughout most of their lives, often are not interested in participating in robot studies from our experience. For example, in Ritschel et al. [2019c,d], people refused to let a robot into their home. People mention privacy concerns irrespective of their age in a variety of cases. It is very important to build social agents in a transparent and privacy respecting manner. However, this limits the technology, which can be used in practice, due to computing power restrictions in the wild, such as speech recognition and natural language processing. Another challenge is that some participants do not have internet or do not permit its usage due to privacy concerns, which prevents loading contents from the internet dynamically. This can be important with regard to the novelty effect, but also for information retrieval tasks or communication.

When investigating machine learning approaches, such as adaptation via reinforcement learning, another challenge arises. Since the agent faces a control problem and needs to decide what to do during runtime, every action manipulates its environment. Consequently, one does not know how the user would have reacted if the agent would have done something else. Especially in the beginning, if the learning agent starts without initial knowledge, this is an inherent problem. When using reinforcement learning, exploration is mandatory. Special care needs to be taken in order to make sure that the agent's randomized action selection does not irritate the human or impact the overall interaction experience. However, variety in the agent's expressed behaviors is essential to keep interaction with the agent interesting over a longer time, since reinforcement learning sticks to the most effective solution most of the time.

In addition, a serious challenge with regard to adaptation in an uncontrolled environment is that human feedback will be biased. Many external aspects may influence the human's reactions, such as the user's mood, stress due to upcoming appointments, but also technical restrictions. For example, when using social signals to calculate the reward signal based on human reactions, misinterpreted sensor data, bad lighting conditions, occlusions or other problems occur, that could even result in not being able to sense human reactions at all. Human feedback can also be biased by the contents presented by the agent, which are not under control of the adaptation process. For example, they may elicit emotions unintentionally and thus influence the feedback towards the agent, which is uncorrelated with the agent's actual actions.

In general, there are also substantial technical challenges when conducting in-situ studies. The agent needs to be completely autonomous and thus, error handling is an important aspect, which becomes increasingly important when several components need to work together. There is no option to control the agent remotely or to make sure that it is working correctly if there is no internet connection on-site. However, autonomous interactive agents (see section 1.4.4) are essential for long-term interaction studies [Leite et al. 2013].

1.6.2 Finding the Right Level of Sensitivity

The further perfectionization of techniques for the analysis of social signals might lead to agents that respond to human signals in an oversensitive manner [Eichner et al. 2007]. Agents, which adapt to human social signals, may irritate users. In case the human was not aware of the expression of his or her own social cues, the agent's reactions might appear random. Obviously, not every social signal cue from the user should trigger a response from the agent. The problem of deciding, which user behavior should be interpreted as system input, is called the Midas Touch Problem. Hoekstra and colleagues [Hoekstra et al. 2007] present a number of strategies to mitigate the Midas Touch Problem for an application with two agents that adapt their presentations to the user's level of attentiveness. In their work, eye gaze was the only user cue that was interpreted by the agents. Thus, the question arises of how to determine the right level of sensitivity for a multitude of social signals in interactive conversational settings.

1.6.3 Fine-Grained Behavior Timing

Personality affects an individual's timing in conversation. For instance, talking over another person is often seen as a sign of dominance [Rogers and Jones 1975]. However, turn-taking is a complex issue and involves many more factors than personality alone [Goldberg 1990]. For instance, completing the speaker's sentence or responding early can also signify attention and engagement, for example when two people are having a passionate conversation about shared interests.

Most systems so far use fixed heuristics, such as yielding the turn immediately upon the user's attempt to interrupt the agent [Bohus and Horvitz 2011], or responding after a fixed time threshold [Visser et al. 2012]. However, computer-controlled characters are faced with an increasing number of different contexts, for example in the case of personal assistants that accompany their user throughout the day and likely for months and years. Therefore, they are in need of turn-taking mechanisms that are appropriate for a wider range of contexts, yet easily adapted to achieve the user-preferred personality impressions.

As brought up in section 1.2.2.3, context-specific behavior preferences can be approached via considering the underlying goals. A system, such as the virtual butler described by Conati [2013], relies on these relationships between internal goals and the personality, which is reflected in a user's or and agent's surface behavior. At the same time, showing appropriate emotional responses to events, that affect a character's goals, is important for making their behavior believable and consequently engaging for the user [op den Akker and Bruijnes 2012].

Decision-theoretic reasoning can then be applied to infer the best course of action based on knowledge of these goals and how the system's actions will affect them [Bohus and Horvitz 2011, Conati 2013, Janowski and André 2018, 2019]. However, identifying the goals, which shape interpersonal timing, is not trivial. While there have been attempts to define the structure behind human motivation [Ortony et al. 1988, Talevich et al. 2017], few works have linked those to a person's Big Five traits or interpersonal stance. Those who do focus on high-level or long-term goals rather than concrete short-term intentions [Reisz et al. 2013]. But in order to make decisions about an agent's behavior, a system needs to reason about short-term goals, such as giving the user time to talk or delivering a message at the earliest opportunity.

1.6.4 Autonomous Interactive Systems

In order to bring socially interactive agents out of the laboratory and into the homes of actual users, it is necessary to verify that the findings from perception studies and Wizard-of-Oz evaluations hold true when autonomous systems have to interact with a human in real-time.

One of the few real-time interaction examples is found in the work of Gebhard et al. [2019]. However, their study relied on a limited domain with pre-scripted dialogue and pre-defined user actions, which is still common practice.

A less constrained setup is described by Skantze et al. [2015]. Their system was exhibited in a museum over nine days, gathering data about turn-taking patterns while pairs of users played a card sorting game with the robot. Though the authors mention the possibility of adjusting the robot's personality through a confidence parameter, that confidence was neither the focus of that study nor was it considered in relation to surface behavior variations.

As mentioned in section 1.4.3, one problem, which makes people turn to Wizard-of-Oz setups or pre-scripted interaction, is the need for quick reaction times. op den Akker and Bruijnes [2012] identified real-time input processing as a major bottleneck in dialogue systems, and a challenge that needs to be solved before one can consider implementing agents that adapt their behavior during conversation. Over the years, there have been different approaches to that. Bohus and Horvitz [2011] explicitly modeled the delays that occur between a user's speech, the system's perception, the selection of a response, and the audible text-to-speech output. Other researchers such as Visser et al. [2012] applied machine learning to try and predict the content of incomplete sentences step by step, in an incremental manner.

While real-time incremental input and output processing form a research topic of their own, it is worth noting that the challenges in this area limit an agent's adaptive behavior with regards to the time frame. This supports the idea that adapting to the slowly-changing, relatively stable personality traits of a user may lead to better results than trying to adapt to, for example, their short-lived emotions.

1.7 Conclusion

Equipping an artificial socially interactive agent with a personality that adapts to the user is a helpful strategy for making the interaction more enjoyable and more engaging. It builds on the humans' natural tendency to anthropomorphize inanimate objects and evaluate their behavior in terms of human codes of conduct. As a consequence, interest in creating believable artificial actors has been growing since the mid-1990ies.

Personality traits are expressed in a number of verbal and nonverbal communication behaviors such as the choice of assertive words, demure gaze behavior or patiently waiting for the right time to respond. By observing these surface cues, both human and machine can infer their conversation partner's disposition for reacting to certain events in a specific manner. This, in turn, enables a socially interactive agent to not only make its internal processes transparent to the user, but also to configure its own personality traits to better suit the user's requirements and personal preferences.

In order to create an adaptive personality, shallow heuristics and superficial mimicking will not suffice in the long run. Future systems need to take the psychological foundations into account if they are to display social competence in a wide range of scenarios. Furthermore, automated learning processes are necessary for tailoring a system to an individual user's preferences while probabilistic reasoning helps to recognize these preferences more accurately in varying contexts. The benefit gained from the agent's action – be it in the form of a reward function or the expected utility – then enables it to choose the optimal strategy for interacting with the user.

Due to the complexities of adaptive systems, its components, such as the altered behavior or a reinforcement learning mechanism, are often evaluated separately. However, to truly understand how to adapt the agent's personality to that of its user, these applications need to be brought out of the laboratories and, ideally, to be tested over an extended period of time with actual members of the target group.

For this purpose, the adaptive agent must meet basic usability standards so that people will actually be motivated to use it long enough for observing the adaptation. This entails that the system must not irritate the user while adapting its behavior. It must be sensitive to the user's feedback, but not too sensitive either. Its actions must be timed properly in different contexts, while still conveying the currently learned personality in a credible manner.

Related research fields, such as incremental input processing and incremental dialog modeling, can provide helpful insights in this regard. Psychological research about human-human communication allows us to understand the mechanisms behind interpersonal compatibility, and how superficial behavior patterns relate to the core personality traits that need to be adapted.

As we learn to equip artificial characters with believable personalities, we also learn about what makes the human mind tick. In the long run, this understanding will allow both human and machine to work together in the most productive, entertaining and supportive manner.

1.8 Acknowledgement

This work has been partially funded by the European Union Horizon 2020 research and innovation programme, grant agreement 856879.

Bibliography

A. Aly and A. Tapus. 2016. Towards an intelligent system for generating an adapted verbal and nonverbal combined behavior in human-robot interaction. *Auton. Robots*, 40(2): 193–209. https://doi.org/10.1007/s10514-015-9444-1. DOI: 10.1007/s10514-015-9444-1.

E. André and C. Pelachaud. 2010. Interacting with embodied conversational agents. In F. Chen and K. Jokinen, eds., *Speech Technology*, pp. 123–149. Springer US. ISBN 978-0-387-73819-2.

E. André, T. Rist, and J. Müller. 1998. Integrating reactive and scripted behaviors in a life-like presentation agent. In K. P. Sycara and M. J. Wooldridge, eds., *Proceedings of the Second International Conference on Autonomous Agents, AGENTS 1998, St. Paul, Minnepolis, USA, May 9-13, 1998*, pp. 261–268. ACM. https://doi.org/10.1145/280765.280842. DOI: 10.1145/280765.280842.

E. André, M. Klesen, P. Gebhard, S. Allen, and T. Rist. 1999. Integrating models of personality and emotions into lifelike characters. In A. Paiva, ed., *Affective Interactions, Towards a New Generation of Computer Interfaces*, volume 1814 of *Lecture Notes in Computer Science*, pp. 150–165. Springer. https://doi.org/10.1007/10720296_11. DOI: 10.1007/10720296_11.

E. André, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. 2000. *The Automated Design of Believable Dialogue*, p. 220–255. MIT Press, Cambridge, MA, USA. ISBN 0262032783.

S. Andrist, B. Mutlu, and A. Tapus. 2015. Look Like Me: Matching Robot Personality via Gaze to Increase Motivation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 3603–3612. ACM, New York, NY, USA. ISBN 978-1-4503-3145-6. http://doi.acm.org/10.1145/2702123.2702592. DOI: 10.1145/2702123.2702592.

D. Arellano, J. Varona, F. J. Perales, N. Bee, K. Janowski, and E. André. 2011. Influence of head orientation in perception of personality traits in virtual agents. In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 3*, AAMAS '11, pp. 1093–1094. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC. ISBN 0-9826571-7-X, 978-0-9826571-7-1. http://dl.acm.org/citation.cfm?id=2034396.2034434.

M. Argyle and M. Cook. 1976. *Gaze and mutual gaze*. Cambridge University Press, The Pitt Building, Trumpington Street, Cambridge CB2 1RP. ISBN 0-521-20865-3.

M. Argyle and B. R. Little. 1972. Do Personality Traits Apply to Social Behaviour? *Journal for the Theory of Social Behaviour*, 2(1): 1–33. ISSN 1468-5914. DOI: 10.1111/j.1468-5914.1972.tb00302.x.

N. I. Badler, B. D. Reich, and B. L. Webber. 1997. Towards personalities for animated agents with reactive and planning behaviors. In P. Petta and R. Trappl, eds., *Creating Personalities for Synthetic Actors, Towards Autonomous Personality Agents*, volume 1195 of *Lecture Notes in Computer Science*, pp. 43–57. Springer. https://doi.org/10.1007/BFb0030569. DOI: 10.1007/BFb0030569.

G. Ball and J. Breese. 2000. *Emotion and Personality in a Conversational Agent*, p. 189–219. MIT Press, Cambridge, MA, USA. ISBN 0262032783.

R. Barraquand and J. L. Crowley. 2008. Learning polite behavior with situation models. In T. Fong, K. Dautenhahn, M. Scheutz, and Y. Demiris, eds., *Proceedings of the 3rd ACM/IEEE international* *conference on Human robot interaction, HRI 2008, Amsterdam, The Netherlands, March 12-15, 2008*, pp. 209–216. ACM. ISBN 978-1-60558-017-3.

N. Bee, S. Franke, and E. Andre. Sept. 2009. Relations between facial display, eye gaze and head tilt: Dominance perception variations of virtual agents. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pp. 1–7. DOI: 10.1109/ACII.2009.5349573.

N. Bee, C. Pollock, E. André, and M. A. Walker. 2010. Bossy or wimpy: Expressing social dominance by combining gaze and linguistic behaviors. In J. M. Allbeck, N. I. Badler, T. W. Bickmore, C. Pelachaud, and A. Safonova, eds., *Intelligent Virtual Agents, 10th International Conference, IVA 2010, Philadelphia, PA, USA, September 20-22, 2010. Proceedings*, volume 6356 of *Lecture Notes in Computer Science*, pp. 265–271. Springer. https://doi.org/10.1007/978-3-642-15892-6_28. DOI: 10.1007/978-3-642-15892-6_28.

E. P. Bernier and B. Scassellati. 2010. The similarity-attraction effect in human-robot interaction. In B. Kuipers, T. R. Shultz, A. Stoytchev, and C. Yu, eds., *2010 IEEE 9th International Conference on Development and Learning, ICDL 2010, Ann Arbor, MI, USA, August 18-21, 2010*, pp. 286–290. IEEE. https://doi.org/10.1109/DEVLRN.2010.5578828. DOI: 10.1109/DEVLRN.2010.5578828.

P. Berry, T. Donneau-Golencer, K. Duong, M. Gervasio, B. Peintner, and N. Yorke-Smith. 2009. Evaluating user-adaptive systems: Lessons from experiences with a personalized meeting scheduling assistant. https://www.aaai.org/ocs/index.php/IAAI/IAAI09/paper/view/271/1014.

C. L. Bethel and R. R. Murphy. 2008. Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Trans. Systems, Man, and Cybernetics, Part C*, 38(1): 83–92. https://doi.org/10.1109/TSMCC.2007.905845. DOI: 10.1109/TSMCC.2007.905845.

T. Bickmore and J. Cassell. 2005. Social dialogue with embodied conversational agents. In *Advances in Natural Multimodal Dialogue Systems*, volume 30 of *Text, Speech and Language Technology*, pp. 23–54. Springer. DOI: 10.1007/1-4020-3933-6_2.

D. Bohus and E. Horvitz. 2011. Decisions about turns in multiparty conversation: From perception to action. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, ICMI '11, pp. 153–160. ACM, New York, NY, USA. ISBN 978-1-4503-0641-6. http://doi.acm.org/10.1145/2070481.2070507. DOI: 10.1145/2070481.2070507.

C. Breazeal. 2004. *Designing sociable robots*. MIT press.

J. Broekens and M. Chetouani. 2019. Towards transparent robot learning through tdrl-based emotional expressions. *IEEE Transactions on Affective Computing*, pp. 1–1. ISSN 1949-3045. DOI: 10.1109/TAFFC.2019.2893348.

P. Brown and S. C. Levinson. Feb. 1987. *Politeness: Some Universals in Language Usage*. Cambridge University Press. ISBN 978-0-521-31355-1.

A. Cafaro, H. H. Vilhjálmsson, T. W. Bickmore, D. Heylen, K. R. Johannsdottir, and G. S. Valgarsson. 2012. First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In Y. I. Nakano, M. Neff, A. Paiva, and M. A. Walker, eds., *Intelligent Virtual Agents - 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings*, volume 7502 of *Lecture Notes in Computer Science*, pp. 67–80. Springer. https://doi.org/10.1007/978-3-642-33197-8_7. DOI: 10.1007/978-3-642-33197-8_7.

A. Cafaro, N. Glas, and C. Pelachaud. 2016. The effects of interrupting behavior on interpersonal attitude and engagement in dyadic interactions. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, AAMAS '16, pp. 911–920. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC. ISBN 978-1-4503-4239-1. http://dl.acm.org/citation.cfm?id=2936924.2937059.

M. Carbonneau, E. Granger, Y. Attabi, and G. Gagnon. 2020. Feature learning from spectrograms for assessment of personality traits. *IEEE Trans. Affective Computing*, 11(1): 25–31. https://doi.org/10.1109/TAFFC.2017.2763132. DOI: 10.1109/TAFFC.2017.2763132.

J. Cassell, C. Pelachaud, N. I. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. 1994. Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In D. Schweitzer, A. S. Glassner, and M. Keeler, eds., *Proceedings of the 21th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1994, Orlando, FL, USA, July 24-29, 1994*, pp. 413–420. ACM. https://doi.org/10.1145/192161.192272. DOI: 10.1145/192161.192272.

J. Cassell, H. H. Vilhjálmsson, and T. W. Bickmore. 2004. BEAT: the behavior expression animation toolkit. In H. Prendinger and M. Ishizuka, eds., *Life-like characters - tools, affective functions, and applications*, Cognitive Technologies, pp. 163–186. Springer.

O. Celiktutan, E. Skordos, and H. Gunes. 2017. Multimodal human-human-robot interactions (mhhri) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing*.

C. Conati. 2013. Virtual Butler: What Can We Learn from Adaptive User Interfaces? In R. Trappl, ed., *Your Virtual Butler*, volume 7407 of *Lecture Notes in Computer Science*, pp. 29–41. Springer, Berlin, Heidelberg. ISBN 978-3-642-37345-9. DOI: 10.1007/978-3-642-37346-6_4.

M. Courgeon and C. Clavel. 2013. MARC: a framework that features emotion models for facial animation during human-computer interaction. *J. Multimodal User Interfaces*, 7(4): 311–319. https://doi.org/10.1007/s12193-013-0124-1. DOI: 10.1007/s12193-013-0124-1.

F. J. F. de Gorostiza Luengo, F. Alonso-Martín, Á. C. González, and M. A. Salichs. 2017. Sound synthesis for communicating nonverbal expressive cues. *IEEE Access*, 5: 1941–1957. https://doi.org/10.1109/ACCESS.2017.2658726. DOI: 10.1109/ACCESS.2017.2658726.

F. den Hengst, E. M. Grua, A. el Hassouni, and M. Hoogendoorn. 2020. Reinforcement learning for personalization: a systematic literature review.

C. G. DeYoung, Y. J. Weisberg, L. C. Quilty, and J. B. Peterson. 2013. Unifying the Aspects of the Big Five, the Interpersonal Circumplex, and Trait Affiliation. *Journal of Personality*, 81(5): 465–475. ISSN 1467-6494. DOI: 10.1111/jopy.12020.

T. Eichner, H. Prendinger, E. André, and M. Ishizuka. 2007. Attentive presentation agents. In C. Pelachaud, J. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé, eds., *Intelligent Virtual Agents, 7th International Conference, IVA 2007, Paris, France, September 17-19, 2007, Proceedings*, volume 4722 of *Lecture Notes in Computer Science*, pp. 283–295. Springer. https://doi.org/10.1007/978-3-540-74997-4_26. DOI: 10.1007/978-3-540-74997-4_26.

S. D. Estroff and S. Nowicki Jr. 1992. Interpersonal Complementarity, Gender of Interactants, and Performance on Puzzle and Word Tasks. *Personality and Social Psychology Bulletin*, 18(3): 351–356. https://doi.org/10.1177/0146167292183012. DOI: 10.1177/0146167292183012. _eprint: https://doi.org/10.1177/0146167292183012.

E. Ferreira and F. Lefèvre. 2015. Reinforcement-learning based dialogue system for human-robot interactions with socially-inspired rewards. *Comput. Speech Lang.*, 34(1): 256–274. https://doi.org/10.1016/j.csl.2015.03.007. DOI: 10.1016/j.csl.2015.03.007.

P. Fournier, O. Sigaud, and M. Chetouani. 2017. Combining artificial curiosity and tutor guidance for environment exploration. In *Workshop on Behavior Adaptation, Interaction and Learning for Assistive Robotics at IEEE RO-MAN 2017*.

P. Gebhard. Nov 2007. *Emotionalisierung interaktiver Virtueller Charaktere. Ein mehrschichtiges Computermodell zur Erzeugung und Simulation von Gefühlen in Echtzeit*. PhD thesis, Universität des Saarlandes, Saarbrücken.

P. Gebhard, T. Schneeberger, G. Mehlmann, T. Baur, and E. André. 2019. Designing the Impression of Social Agents' Real-time Interruption Handling. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, IVA '19, pp. 19–21. ACM, New York, NY, USA. ISBN 978-1-4503-6672-4. http://doi.acm.org/10.1145/3308532.3329435. DOI: 10.1145/3308532.3329435. event-place: Paris, France.

J. A. Goldberg. 1990. Interrupting the discourse on interruptions: An analysis in terms of relationally neutral, power- and rapport-oriented acts. *Journal of Pragmatics*, 14(6): 883 – 903. ISSN 0378-2166. DOI: 10.1016/0378-2166(90)90045-F.

G. Gordon, S. Spaulding, J. K. Westlund, J. J. Lee, L. Plummer, M. Martinez, M. Das, and C. Breazeal. 2016. Affective personalization of a social robot tutor for children's second language skills. In D. Schuurmans and M. P. Wellman, eds., *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.*, pp. 3951–3957. AAAI Press. ISBN 978-1-57735-760-5. http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11759.

J. Gratch, J. Rickel, E. André, J. Cassell, E. Petajan, and N. I. Badler. 2002. Creating interactive virtual humans: Some assembly required. *IEEE Intell. Syst.*, 17(4): 54–63. https://doi.org/10.1109/MIS.2002.1024753. DOI: 10.1109/MIS.2002.1024753.

S. Hammer, B. Lugrin, S. Bogomolov, K. Janowski, and E. André. 2016. Investigating politeness strategies and their persuasiveness for a robotic elderly assistant. In A. Meschtscherjakov, B. E. R. de Ruyter, V. Fuchsberger, M. Murer, and M. Tscheligi, eds., *Persuasive Technology - 11th International Conference, PERSUASIVE 2016, Salzburg, Austria, April 5-7, 2016, Proceedings*, volume 9638 of *Lecture Notes in Computer Science*, pp. 315–326. Springer. https://doi.org/10.1007/978-3-319-31510-2_27. DOI: 10.1007/978-3-319-31510-2_27.

V. Harrison, L. Reed, S. Oraby, and M. A. Walker. 2019. Maximizing stylistic control and semantic accuracy in NLG: personality variation and discourse contrast. *CoRR*, abs/1907.09527. http://arxiv.org/abs/1907.09527.

B. Hartmann, M. Mancini, and C. Pelachaud. 2005. Implementing expressive gesture synthesis for embodied conversational agents. In S. Gibet, N. Courty, and J. Kamp, eds., *Gesture in Human-Computer Interaction and Simulation, 6th International Gesture Workshop, GW 2005, Berder Island, France, May 18-20, 2005, Revised Selected Papers*, volume 3881 of *Lecture Notes in Computer Science*, pp. 188–199. Springer. https://doi.org/10.1007/11678816_22. DOI: 10.1007/11678816_22.

K. Hayashi, T. Kanda, T. Miyashita, H. Ishiguro, and N. Hagita. 2008. Robot *Manzai*: Robot conversation as a passive-social medium. *I. J. Humanoid Robotics*, 5(1): 67–86. https://doi.org/10.1142/S0219843608001315. DOI: 10.1142/S0219843608001315.

B. Hayes-Roth, R. van Gent, and D. Huber. 1997. Acting in character. In P. Petta and R. Trappl, eds., *Creating Personalities for Synthetic Actors, Towards Autonomous Personality Agents*, volume 1195 of *Lecture Notes in Computer Science*, pp. 92–112. Springer. https://doi.org/10.1007/BFb0030573. DOI: 10.1007/BFb0030573.

J. Hemminghaus and S. Kopp. 2017. Towards adaptive social behavior generation for assistive robots using reinforcement learning. In B. Mutlu, M. Tscheligi, A. Weiss, and J. E. Young, eds., *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, HRI 2017, Vienna, Austria, March 6-9, 2017*, pp. 332–340. ACM. ISBN 978-1-4503-4336-7. http://dl.acm.org/citation.cfm?id=3020217. DOI: 10.1145/2909824.

R. Hoegen, D. Aneja, D. J. McDuff, and M. Czerwinski. 2019. An end-to-end conversational style matching agent. In C. Pelachaud, J. Martin, H. Buschmeier, G. M. Lucas, and S. Kopp, eds., *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, IVA 2019, Paris, France, July 2-5, 2019*, pp. 111–118. ACM. https://doi.org/10.1145/3308532.3329473. DOI: 10.1145/3308532.3329473.

A. Hoekstra, H. Prendinger, N. Bee, D. Heylen, and M. Ishizuka. 2007. Highly realistic 3d presentation agents with visual attention capability. In A. Butz, B. D. Fisher, A. Krüger, P. Olivier, and S. Owada, eds., *Smart Graphics, 7th International Symposium, SG 2007, Kyoto, Japan, June 25-27, 2007, Proceedings*, volume 4569 of *Lecture Notes in Computer Science*, pp. 73–84. Springer. https://doi.org/10.1007/978-3-540-73214-3_7. DOI: 10.1007/978-3-540-73214-3_7.

L. M. Horowitz, K. R. Wilson, B. Turan, P. Zolotsev, M. J. Constantino, and L. Henderson. 2006. How Interpersonal Motives Clarify the Meaning of Interpersonal Behavior: A Revised Circumplex Model. *Personality and Social Psychology Review*, 10(1): 67–86. DOI: 10.1207/s15327957pspr1001_4.

C. Hu, M. A. Walker, M. Neff, and J. E. F. Tree. 2015. Storytelling agents with personality and adaptivity. In W. Brinkman, J. Broekens, and D. Heylen, eds., *Intelligent Virtual Agents - 15th International Conference, IVA 2015, Delft, The Netherlands, August 26-28, 2015, Proceedings*, volume 9238 of *Lecture Notes in Computer Science*, pp. 181–193. Springer. https://doi.org/10.1007/978-3-319-21996-7_19. DOI: 10.1007/978-3-319-21996-7_19.

Z. Hu, J. E. F. Tree, and M. A. Walker. 2018. Modeling linguistic and personality adaptation for natural language generation. In K. Komatani, D. J. Litman, K. Yu, L. Cavedon, M. Nakano, and A. Papangelis, eds., *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue, Melbourne, Australia, July 12-14, 2018*, pp. 20–31. Association for Computational Linguistics. https://doi.org/10.18653/v1/w18-5003. DOI: 10.18653/v1/w18-5003.

F. Iida, M. Tabata, and F. Hara. 1998. Generating personality character in a face robot through interaction with human. In *7th IEEE International Workshop on Robot and Human Communication*, pp. 481–486.

K. Isbister and C. Nass. 2000. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *Int. J. Hum. Comput. Stud.*, 53(2): 251–267. https://doi.org/10.1006/ijhc.2000.0368. DOI: 10.1006/ijhc.2000.0368.

A. Jain, F. Pecune, Y. Matsuyama, and J. Cassell. 2018. A user simulator architecture for socially-aware conversational agents. In A. Bogdanovych, D. Richards, S. Simoff, C. Pelachaud, D. Heylen, and T. Trescak, eds., *Proceedings of the 18th International Conference on Intelligent Virtual Agents, IVA 2018, Sydney, NSW, Australia, November 05-08, 2018*, pp. 133–140. ACM. https://doi.org/10.1145/3267851.3267916. DOI: 10.1145/3267851.3267916.

K. Janowski and E. André. 2018. Decision-theoretic personality-based reasoning about turn-taking conflicts. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, IVA '18, pp. 349–350. ACM, New York, NY, USA. ISBN 978-1-4503-6013-5. http://doi.acm.org/10.1145/3267851.3267899. DOI: 10.1145/3267851.3267899.

K. Janowski and E. André. 2019. What If I Speak Now? A Decision-Theoretic Approach to Personality-Based Turn-Taking. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '19, pp. 1051–1059. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC. ISBN 978-1-4503-6309-9.

K. Janowski, H. Ritschel, B. Lugrin, and E. Andre. 2018. Sozial interagierende roboter in der pflege. In O. Bendel, ed., *Pflegeroboter*. ISBN 9783658226978. DOI: 10.1007/978-3-658-22698-5_4.

W. L. Johnson, R. E. Mayer, E. André, and M. Rehm. 2005. Cross-cultural evaluation of politeness in tactics for pedagogical agents. In C. Looi, G. I. McCalla, B. Bredeweg, and J. Breuker, eds., *Artificial Intelligence in Education - Supporting Learning through Intelligent and Socially Informed Technology, Proceedings of the 12th International Conference on Artificial Intelligence in Education, AIED 2005, July 18-22, 2005, Amsterdam, The Netherlands*, volume 125 of *Frontiers in Artificial Intelligence and Applications*, pp. 298–305. IOS Press. ISBN 978-1-58603-530-3. http://www.booksonline.iospress.nl/Content/View.aspx?piid=1322.

M. Joosse, M. Lohse, J. G. Perez, and V. Evers. 2013. What you do is who you are: The role of task context in perceived social robot personality. In *2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, May 6-10, 2013*, pp. 2134–2139. IEEE. https://doi.org/10.1109/ICRA.2013.6630863. DOI: 10.1109/ICRA.2013.6630863.

K. Katevas, P. G. Healey, and M. T. Harris. 2015. Robot comedy lab: experimenting with the social dynamics of live performance. *Frontiers in psychology*, 6.

E. S. Kim and B. Scassellati. July 2007. Learning to refine behavior using prosodic feedback. In *2007 IEEE 6th International Conference on Development and Learning*, pp. 205–210. DOI: 10.1109/DEVLRN.2007.4354072.

H. Kim, S. S. Kwak, and M. Kim. 2008. Personality design of sociable robots by control of gesture design factors. In M. Buss and K. Kühnlenz, eds., *The 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2008, Munich, Germany, August 1-3, 2008*, pp. 494–499. IEEE. https://doi.org/10.1109/ROMAN.2008.4600715. DOI: 10.1109/ROMAN.2008.4600715.

M. L. Knapp, J. A. Hall, and T. G. Horgan. 2014. *Nonverbal Communication in Human Interaction (International Edition)*, 8. Wadsworth, Cengage Learning. ISBN 1-285-08351-2.

H. Knight. 2011. Eight lessons learned about non-verbal interactions through robot theater. In B. Mutlu, C. Bartneck, J. Ham, V. Evers, and T. Kanda, eds., *Social Robotics - Third International Conference, ICSR 2011, Amsterdam, The Netherlands, November 24-25, 2011. Proceedings*, volume 7072 of *Lecture Notes in Computer Science*, pp. 42–51. Springer. ISBN 978-3-642-25503-8. https://doi.org/10.1007/978-3-642-25504-5_5. DOI: 10.1007/978-3-642-25504-5_5.

T. Kucherenko, P. Jonell, S. van Waveren, G. Eje Henter, S. Alexanderson, I. Leite, and H. Kjellström. 2020. Gesticulator: A framework for semantically-aware speech-driven gesture generation. *arXiv*, pp. arXiv–2001.

B. Laurel. 1997. *Interface Agents: Metaphors with Character*, p. 67–77. MIT Press, Cambridge, MA, USA. ISBN 0262522349.

I. Leite, A. Pereira, G. Castellano, S. Mascarenhas, C. Martinho, and A. Paiva. 2011. Modelling empathy in social robotic companions. In L. Ardissono and T. Kuflik, eds., *Advances in User Modeling - UMAP 2011 Workshops, Girona, Spain, July 11-15, 2011, Revised Selected Papers*, volume 7138 of *Lecture Notes in Computer Science*, pp. 135–147. Springer. ISBN 978-3-642-28508-0. https://doi.org/10.1007/978-3-642-28509-7_14. DOI: 10.1007/978-3-642-28509-7$_1$4.

I. Leite, C. Martinho, and A. Paiva. 2013. Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, 5(2): 291–308.

T. W. Liew and S. Tan. 2016. Virtual agents with personality: Adaptation of learner-agent personality in a virtual learning environment. In *Eleventh International Conference on Digital Information Management, ICDIM 2016, Porto, Portugal, September 19-21, 2016*, pp. 157–162. IEEE. https://doi.org/10.1109/ICDIM.2016.7829758. DOI: 10.1109/ICDIM.2016.7829758.

C. Liu, K. Conn, N. Sarkar, and W. Stone. 2008. Online affect detection and robot behavior adaptation for intervention of children with autism. *IEEE Trans. Robotics*, 24(4): 883–896. https://doi.org/10.1109/TRO.2008.2001362. DOI: 10.1109/TRO.2008.2001362.

F. Mairesse and M. A. Walker. 2010. Towards personality-based user adaptation: psychologically informed stylistic language generation. *User Model. User-Adapt. Interact.*, 20(3): 227–278. https://doi.org/10.1007/s11257-010-9076-2. DOI: 10.1007/s11257-010-9076-2.

F. Makedon, ed. 2019. *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA 2019, Island of Rhodes, Greece, June 5-7, 2019*. ACM. ISBN 978-1-4503-6232-0. https://doi.org/10.1145/3316782. DOI: 10.1145/3316782.

M. Mancini, B. Biancardi, S. Dermouche, P. Lerner, and C. Pelachaud. 2019. Managing agent's impression based on user's engagement detection. In C. Pelachaud, J. Martin, H. Buschmeier, G. M. Lucas, and S. Kopp, eds., *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, IVA 2019, Paris, France, July 2-5, 2019*, pp. 209–211. ACM. https://doi.org/10.1145/3308532.3329442. DOI: 10.1145/3308532.3329442.

P. M. Markey, D. C. Funder, and D. J. Ozer. 2003. Complementarity of Interpersonal Behaviors in Dyadic Interactions. *Personality and Social Psychology Bulletin*, 29(9): 1082–1090. https://doi.org/10.1177/0146167203253474. DOI: 10.1177/0146167203253474. _eprint: https://doi.org/10.1177/0146167203253474.

G. S. Martins, L. Santos, and J. Dias. Jun 2018. User-adaptive interaction in social robots: A survey focusing on non-physical interaction. *International Journal of Social Robotics*. ISSN 1875-4805. https://doi.org/10.1007/s12369-018-0485-4. DOI: 10.1007/s12369-018-0485-4.

R. McCrae and O. John. 1992. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2): 175.

R. R. McCrae and P. T. Costa. 1989. The structure of interpersonal traits: Wiggins's circumplex and the five-factor model. *Journal of personality and social psychology*, 56(4): 586.

A. Mehrabian. Dec. 1996a. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology*, 14(4): 261–292. ISSN 1936-4733. https://doi.org/10.1007/BF02686918. DOI: 10.1007/BF02686918.

A. Mehrabian. Aug. 1996b. Analysis of the Big-five Personality Factors in Terms of the PAD Temperament Model. *Australian Journal of Psychology*, 48(2): 86–92. ISSN 00049530. DOI: 10.1080/00049539608259510.

N. Mitsunaga, C. Smith, T. Kanda, H. Ishiguro, and N. Hagita. 2008. Adapting robot behavior for human–robot interaction. *IEEE Trans. Robotics*, 24(4): 911–916. https://doi.org/10.1109/TRO.2008.926867. DOI: 10.1109/TRO.2008.926867.

R. Montoya and R. Horton. 02 2013. A meta-analytic investigation of the processes underlying the similarity-attraction effect. *Journal of Social and Personal Relationships*, 30: 64–94. DOI: 10.1177/0265407512452989.

Y. Moon. 2002. Personalization and Personality: Some Effects of Customizing Message Style Based on Consumer Personality. *Journal of Consumer Psychology*, 12(4): 313 – 325. ISSN 1057-7408. http://www.sciencedirect.com/science/article/pii/S1057740816300833. DOI: https://doi.org/10.1016/S1057-7408(16)30083-3.

Y. Moon and C. I. Nass. 1996. Adaptive agents and personality change: Complementarity versus similarity as forms of adaptation. In *Conference Companion on Human Factors in Computing Systems*, CHI '96, p. 287–288. Association for Computing Machinery, New York, NY, USA. ISBN 0897918320. https://doi.org/10.1145/257089.257325. DOI: 10.1145/257089.257325.

S. J. Mountford, B. Buxton, M. W. Krueger, B. Laurel, and L. Vertelney. 1989. Drama and personality in user interface design. In K. Bice and C. H. Lewis, eds., *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1989, Austin, Texas, USA, April 30 - June 4, 1989*, pp. 105–108. ACM. https://doi.org/10.1145/67449.67471. DOI: 10.1145/67449.67471.

A. Najar, O. Sigaud, and M. Chetouani. 2016. Training a robot with evaluative feedback and unlabeled guidance signals. In *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016, New York, NY, USA, August 26-31, 2016*, pp. 261–266. IEEE. ISBN 978-1-5090-3929-6. https://doi.org/10.1109/ROMAN.2016.7745140. DOI: 10.1109/ROMAN.2016.7745140.

C. Nass, Y. Moon, B. J. Fogg, B. Reeves, and D. C. Dryer. 1995. Can computer personalities be human personalities? *Int. J. Hum. Comput. Stud.*, 43(2): 223–239. https://doi.org/10.1006/ijhc.1995.1042. DOI: 10.1006/ijhc.1995.1042.

M. Neff, Y. Wang, R. Abbott, and M. A. Walker. 2010. Evaluating the effect of gesture and language on personality perception in conversational agents. In J. M. Allbeck, N. I. Badler, T. W. Bickmore, C. Pelachaud, and A. Safonova, eds., *Intelligent Virtual Agents, 10th International Conference, IVA 2010, Philadelphia, PA, USA, September 20-22, 2010. Proceedings*, volume 6356 of *Lecture Notes in Computer Science*, pp. 222–235. Springer. https://doi.org/10.1007/978-3-642-15892-6_24. DOI: 10.1007/978-3-642-15892-6_24.

H. Nguyen, D. Morales, and T. Chin. 2018. A neural chatbot with personality. Technical report, Stanford University.

J. Oakman, S. Gifford, and N. Chlebowsky. 2003. A Multilevel Analysis of the Interpersonal Behavior of Socially Anxious People. *Journal of Personality*, 71(3): 397–434. https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-6494.7103006. DOI: 10.1111/1467-6494.7103006.

C. Oertel, G. Castellano, M. Chetouani, J. Nasir, M. Obaid, C. Pelachaud, and C. E. Peters. 2020. Engagement in human-agent interaction: An overview. *Frontiers in Robotics and AI*, 7: 92.

R. op den Akker and M. Bruijnes. 2012. Computational models of social and emotional turn-taking for embodied conversational agents: a review. *COMMIT deliverable*.

S. Oraby, L. Reed, S. Tandon, S. T. S., S. M. Lukin, and M. A. Walker. 2018. Controlling personality-based stylistic variation with neural natural language generators. In K. Komatani, D. J. Litman, K. Yu,

L. Cavedon, M. Nakano, and A. Papangelis, eds., *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue, Melbourne, Australia, July 12-14, 2018*, pp. 180–190. Association for Computational Linguistics. https://doi.org/10.18653/v1/w18-5019. DOI: 10.18653/v1/w18-5019.

A. Ortony, G. L. Clore, and A. Collins. 1988. *The Cognitive Structure of Emotions*. Cambridge University Press. ISBN 978-0-511-57129-9. http://dx.doi.org/10.1017/CBO9780511571299.

P. Petta and R. Trappl, eds. 1997. *Creating Personalities for Synthetic Actors, Towards Autonomous Personality Agents*, volume 1195 of *Lecture Notes in Computer Science*. Springer. ISBN 3-540-62735-9. https://doi.org/10.1007/BFb0030565. DOI: 10.1007/BFb0030565.

O. Pietquin. 2013. Inverse reinforcement learning for interactive systems. In H. Cuayáhuitl, L. Frommberger, N. Dethlefs, and M. van Otterlo, eds., *Proceedings of the 2nd Workshop on Machine Learning for Interactive Systems - Bridging the Gap Between Perception, Action and Communication, MLIS@IJCAI 2013, Beijing, China, August 4, 2013*, pp. 71–75. ACM. https://doi.org/10.1145/2493525.2493529. DOI: 10.1145/2493525.2493529.

B. Ravenet, A. Cafaro, B. Biancardi, M. Ochs, and C. Pelachaud. 2015. Conversational Behavior Reflecting Interpersonal Attitudes in Small Group Interactions. In W.-P. Brinkman, J. Broekens, and D. Heylen, eds., *Intelligent Virtual Agents*, volume 9238, pp. 375–388. Springer International Publishing, Cham. ISBN 978-3-319-21995-0 978-3-319-21996-7. DOI: 10.1007/978-3-319-21996-7_41.

B. Reeves and C. Nass. 1996a. *The media equation - how people treat computers, television, and new media like real people and places*. Cambridge University Press. ISBN 978-1-57586-052-7.

B. Reeves and C. Nass. 1996b. *The Media Equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press, New York, NY, USA.

Z. Reisz, M. Boudreaux, and D. Ozer. 2013. Personality traits and the prediction of personal goals. *Personality and Individual Differences*, 55. DOI: 10.1016/j.paid.2013.05.023.

V. Rieser and O. Lemon. 2011. *Reinforcement Learning for Adaptive Dialogue Systems - A Data-driven Methodology for Dialogue Management and Natural Language Generation*. Theory and Applications of Natural Language Processing. Springer. ISBN 978-3-642-24941-9. https://doi.org/10.1007/978-3-642-24942-6. DOI: 10.1007/978-3-642-24942-6.

T. Rist, E. André, and S. Baldes. 2003. A flexible platform for building applications with life-like characters. In D. B. Leake, W. L. Johnson, and E. André, eds., *Proceedings of the 8th International Conference on Intelligent User Interfaces, IUI 2003, Miami, FL, USA, January 12-15, 2003*, pp. 158–165. ACM. https://doi.org/10.1145/604045.604071. DOI: 10.1145/604045.604071.

T. Rist, E. André, S. Baldes, P. Gebhard, M. Klesen, M. Kipp, P. Rist, and M. Schmitt. 2004. A review of the development of embodied presentation agents and their application fields. In H. Prendinger and M. Ishizuka, eds., *Life-like characters - tools, affective functions, and applications*, Cognitive Technologies, pp. 377–404. Springer.

H. Ritschel. 2018. Socially-aware reinforcement learning for personalized human-robot interaction. In E. André, S. Koenig, M. Dastani, and G. Sukthankar, eds., *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2018, Stockholm, Sweden, July 10-15, 2018*, pp. 1775–1777. International Foundation for Autonomous Agents and Multiagent Systems Richland, SC, USA / ACM. http://dl.acm.org/citation.cfm?id=3237972.

H. Ritschel and E. André. 2017. Real-time robot personality adaptation based on reinforcement learning and social signals. In B. Mutlu, M. Tscheligi, A. Weiss, and J. E. Young, eds., *Companion* *of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, HRI 2017, Vienna, Austria, March 6-9, 2017*, pp. 265–266. ACM. https://doi.org/10.1145/3029798.3038381. DOI: 10.1145/3029798.3038381.

H. Ritschel and E. André. 2018. Shaping a social robot's humor with natural language generation and socially-aware reinforcement learning. In *Proceedings of the Workshop on NLG for Human–Robot Interaction*, pp. 12–16.

H. Ritschel, T. Baur, and E. André. 2017. Adapting a robot's linguistic style based on socially-aware reinforcement learning. In *26th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2017, Lisbon, Portugal, August 28 - Sept. 1, 2017*, pp. 378–384. IEEE. ISBN 978-1-5386-3518-6. https://doi.org/10.1109/ROMAN.2017.8172330. DOI: 10.1109/ROMAN.2017.8172330.

H. Ritschel, A. Seiderer, K. Janowski, I. Aslan, and E. André. 2018. Drink-o-mender: An adaptive robotic drink adviser. In *Proceedings of the 3rd International Workshop on Multisensory Approaches to Human-Food Interaction*, MHFI'18, pp. 3:1–3:8. ACM, New York, NY, USA. ISBN 978-1-4503-6074-6. http://doi.acm.org/10.1145/3279954.3279957. DOI: 10.1145/3279954.3279957.

H. Ritschel, I. Aslan, S. Mertes, A. Seiderer, and E. André. 2019a. Personalized synthesis of intentional and emotional non-verbal sounds for social robots. In *8th International Conference on Affective Computing and Intelligent Interaction, ACII 2019, Cambridge, United Kingdom, September 3-6, 2019*, pp. 1–7. IEEE. https://doi.org/10.1109/ACII.2019.8925487. DOI: 10.1109/ACII.2019.8925487.

H. Ritschel, K. Janowski, A. Seiderer, and E. André. 2019b. Towards a robotic dietitian with adaptive linguistic style. In E. C. Strinati, D. Charitos, I. Chatzigiannakis, P. Ciampolini, F. Cuomo, P. D. Lorenzo, D. Gavalas, S. Hanke, A. Komninos, and G. Mylonas, eds., *Joint Proceeding of the Poster and Workshop Sessions of AmI-2019, the 2019 European Conference on Ambient Intelligence, Rome, Italy, November 13-15, 2019*, volume 2492 of *CEUR Workshop Proceedings*, pp. 134–138. CEUR-WS.org. http://ceur-ws.org/Vol-2492/paper16.pdf.

H. Ritschel, K. Janowski, A. Seiderer, S. Wagner, and E. André. 2019c. Insights on usability and user feedback for an assistive robotic health companion with adaptive linguistic style. In Makedon [2019], pp. 319–320. ISBN 978-1-4503-6232-0. https://doi.org/10.1145/3316782.3322737. DOI: 10.1145/3316782.3322737.

H. Ritschel, A. Seiderer, K. Janowski, S. Wagner, and E. André. 2019d. Adaptive linguistic style for an assistive robotic health companion based on explicit human feedback. In Makedon [2019], pp. 247–255. ISBN 978-1-4503-6232-0. https://doi.org/10.1145/3316782.3316791. DOI: 10.1145/3316782.3316791.

H. Ritschel, T. Kiderle, K. Weber, F. Lingenfelser, T. Baur, and E. André. 2020a. Multimodal joke generation and paralinguistic personalization for a socially-aware robot. In Y. Demazeau, T. Holvoet, J. M. Corchado, and S. Costantini, eds., *Advances in Practical Applications of Agents, Multi-Agent Systems, and Trustworthiness. The PAAMS Collection*, pp. 278–290. Springer International Publishing, Cham. ISBN 978-3-030-49778-1.

H. Ritschel, A. Seiderer, and E. André. 2020b. Pianobot: An adaptive robotic piano tutor. In *Workshop on Exploring Creative Content in Social Robotics at HRI 2020*. https://mypersonalrobots.org/s/HRI_2020_WS_Creative_SR_paper_10.pdf.

L. P. Robert Jr., R. Alahmad, C. Esterwood, S. Kim, S. You, and Q. Zhang. 2020. A review of personality in human–robot interactions. *Foundations and Trends® in Information Systems*, 4(2): 107–212. ISSN 2331-1231. http://dx.doi.org/10.1561/2900000018. DOI: 10.1561/2900000018.

W. T. Rogers and S. S. Jones. 1975. Effects Of Dominance Tendencies On Floor Holding And Interruption Behavior In Dyadic Interaction. *Human Communication Research*, 1(2): 113–122. DOI: 10.1111/j.1468-2958.1975.tb00259.x.

H. Salam, O. Çeliktutan, I. H. Torres, H. Gunes, and M. Chetouani. 2017. Fully automatic analysis of engagement and its relationship to personality in human-robot interactions. *IEEE Access*, 5: 705–721. https://doi.org/10.1109/ACCESS.2016.2614525. DOI: 10.1109/ACCESS.2016.2614525.

D. Schiller, K. Weitz, K. Janowski, and E. André. 2019. Human-inspired socially-aware interfaces. In C. Martín-Vide, G. T. Pond, and M. A. Vega-Rodríguez, eds., *Theory and Practice of Natural Computing - 8th International Conference, TPNC 2019, Kingston, ON, Canada, December 9-11, 2019, Proceedings*, volume 11934 of *Lecture Notes in Computer Science*, pp. 41–53. Springer. https://doi.org/10.1007/978-3-030-34500-6_2. DOI: 10.1007/978-3-030-34500-6_2.

S. Schneider and F. Kummert. 2020. Comparing robot and human guided personalization: Adaptive exercise robots are perceived as more competent and trustworthy. *International Journal of Social Robotics*, pp. 1–17.

G. Skantze, M. Johansson, and J. Beskow. 2015. Exploring turn-taking cues in multi-party human-robot discussions about objects. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, p. 67–74. Association for Computing Machinery, New York, NY, USA. ISBN 9781450339124. https://doi.org/10.1145/2818346.2820749. DOI: 10.1145/2818346.2820749.

H. J. Smith, C. Cao, M. Neff, and Y. Wang. 2019. Efficient neural networks for real-time motion style transfer. *Proc. ACM Comput. Graph. Interact. Tech.*, 2(2): 13:1–13:17. https://doi.org/10.1145/3340254. DOI: 10.1145/3340254.

R. S. Sutton and A. G. Barto. 2018. *Reinforcement Learning - An Introduction (Second Edition)*. Adaptive Computation and Machine Learning. MIT Press. ISBN 9780262039246.

J. R. Talevich, S. J. Read, D. A. Walsh, R. Iyer, and G. Chopra. Feb. 2017. Toward a comprehensive taxonomy of human motives. *PloS one*, 12(2). ISSN 1932-6203. https://www.ncbi.nlm.nih.gov/pubmed/28231252. DOI: 10.1371/journal.pone.0172279.

A. Tapus, C. Tapus, and M. J. Mataric. 2008. User - robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics*, 1(2): 169–183. https://doi.org/10.1007/s11370-008-0017-4. DOI: 10.1007/s11370-008-0017-4.

M. ter Maat, K. P. Truong, and D. K. J. Heylen. Oct. 2011. How Agents' Turn-Taking Strategies Influence Impressions and Response Behaviors. *Presence: Teleoperators and Virtual Environments*, 20(5): 412–430. ISSN 1054-7460. DOI: 10.1162/PRES_a_00064.

R. P. Tett and P. J. Murphy. Dec. 2002. Personality and Situations in Co-worker Preference: Similarity and Complementarity in Worker Compatibility. *Journal of Business and Psychology*, 17(2): 223–243. ISSN 1573-353X. https://doi.org/10.1023/A:1019685515745. DOI: 10.1023/A:1019685515745.

M. Thiébaux, S. Marsella, A. N. Marshall, and M. Kallmann. 2008. Smartbody: behavior realization for embodied conversational agents. In L. Padgham, D. C. Parkes, J. P. Müller, and S. Parsons, eds., *7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008), Estoril, Portugal, May 12-16, 2008, Volume 1*, pp. 151–158. IFAAMAS. https://dl.acm.org/citation.cfm?id=1402409.

K. Tsiakas, M. Abujelala, and F. Makedon. 2018. Task engagement as personalization feedback for socially-assistive robots and cognitive training. *Technologies*, 6(2). ISSN 2227-7080. http://www.mdpi.

com/2227-7080/6/2/49. DOI: 10.3390/technologies6020049.

H. H. Vilhjálmsson, N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. R. Thórisson, H. van Welbergen, and R. J. van der Werf. 2007. The behavior markup language: Recent developments and challenges. In C. Pelachaud, J. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé, eds., *Intelligent Virtual Agents, 7th International Conference, IVA 2007, Paris, France, September 17-19, 2007, Proceedings*, volume 4722 of *Lecture Notes in Computer Science*, pp. 99–111. Springer. https://doi.org/10.1007/978-3-540-74997-4_10. DOI: 10.1007/978-3-540-74997-4_10.

A. Vinciarelli and G. Mohammadi. 2014. A survey of personality computing. *IEEE Trans. Affective Computing*, 5(3): 273–291. https://doi.org/10.1109/TAFFC.2014.2330816. DOI: 10.1109/TAFFC.2014.2330816.

T. Visser, D. Traum, D. DeVault, and R. op den Akker. 2012. Toward a model for incremental grounding in spoken dialogue systems. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents*.

K. Wada and T. Shibata. 2006. Robot therapy in a care house - its sociopsychological and physiological effects on the residents. In *Proceedings of the 2006 IEEE International Conference on Robotics and Automation, ICRA 2006, May 15-19, 2006, Orlando, Florida, USA*, pp. 3966–3971. IEEE. https://doi.org/10.1109/ROBOT.2006.1642310. DOI: 10.1109/ROBOT.2006.1642310.

M. A. Walker, J. E. Cahn, and S. Whittaker. 1997. Improvising linguistic style: Social and affective bases for agents. In W. L. Johnson, ed., *Proceedings of the First International Conference on Autonomous Agents, AGENTS 1997, Marina del Rey, California, USA, February 5-8, 1997*, pp. 96–105. ACM. https://doi.org/10.1145/267658.267680. DOI: 10.1145/267658.267680.

K. Weber, H. Ritschel, I. Aslan, F. Lingenfelser, and E. André. 2018. How to shape the humor of a robot - social behavior adaptation based on reinforcement learning. In S. K. D'Mello, P. G. Georgiou, S. Scherer, E. M. Provost, M. Soleymani, and M. Worsley, eds., *Proceedings of the 2018 on International Conference on Multimodal Interaction, ICMI 2018, Boulder, CO, USA, October 16-20, 2018*, pp. 154–162. ACM. http://doi.acm.org/10.1145/3242969.3242976. DOI: 10.1145/3242969.3242976.