Two Techniques for Assessing Virtual Agent Personality

Kris Liu, Jackson Tolins, Jean E. Fox Tree, Michael Neff and Marilyn A. Walker

Abstract— Personality can be assessed with standardized inventory questions with scaled responses such as "How extraverted is this character?" or with open-ended questions assessing first impressions, such as "What personality does this character convey?" Little is known about how the two methods compare to each other, and even less is known about their use in the personality assessment of virtual agents. We tested what personality virtual agents conveyed through gesture alone when the agents were programmed to display introversion versus extraversion (Experiment 1) and high versus low emotional stability (Experiment 2). In Experiment 1, both measures indicated participants perceived the extraverted agent as extraverted, but the open-question technique highlighted the perception of both agents as highly agreeable whereas the inventory indicated that the extraverted agents were also perceived as more open to new experiences. In Experiment 2, participants perceived agents expressing high versus low emotional stability differently depending on assessment style. With inventory questions, the agents differed on both emotional stability and agreeableness. With the open-ended question, participants perceived the high stability agent as extraverted and the low stability agent as disagreeable. Inventory and open-ended questions provide different information about what personality virtual agents convey and both may be useful in agent development.

Index Terms— Animations, Evaluation/Methodology, Psychology, Social and Behavioral Sciences

——————————————— ◆ ———————————————

1 INTRODUCTION

WHEN observing other humans, people often use subtle behavioral and linguistic cues to make quick determinations of the type of person they are dealing with [1-4]. Increasingly sophisticated virtual agents are now sometimes designed to mimic these cues in order to induce human observers to make similar attributions about the agent with whom they are interacting. The generation of consistent, human-like patterns of expression based on prescribed personality traits makes interaction with these agents easier for users [5, 6] and similarly allows for the beneficial tailoring of agents to specific users or to specific task domains [7-14]. Furthermore, there is evidence that suggests that these cues are most effective when cues are congruent, rather than conflicting [12, 15] .

Personality in virtual agents can be expressed through a number of different cues, such as language usage [13, 16-18], facial expression [6, 8, 19-21], and non-verbal behaviors such as gesture and posture [6, 18, 22, 23]. For example, extraversion is associated with a faster rate of speech and larger gestures, so agents portraying

extraversion talk faster and gesture with wider arms than agents portraying introversion. That personality can be expressed through multiple cues introduces two related yet distinct tasks: designing specific behaviors within a certain modality (i.e., speech or movement) that users will associate readily with a desired personality trait and choosing combinations of behaviors across modalities that are internally consistent with a target personality trait. In other words, behaviors meant to cue specific traits should be accurately assessed in isolation and then behaviors that all cue the same trait can be combined (and assessed). This allows for a consistent portrayal that easily conveys the desired personality notes.

Both of these tasks require accurate assessment of the agent's perceived personality. Personality attribution is generally assessed using standardized Big Five personality questionnaires [13, 16, 22, 24-26], such as the Ten-Item Personality Inventory (TIPI) [27] or Saucier's mini-markers [28]. Yet, determining the best methods for evaluating the behaviors of virtual agents is still an open area of research [29]. So far, relatively few studies assess whether Big Five personality questionnaires, which were designed to measure human personality, are reliable when they are applied to virtual agents [16, 19].

In the current study, we contrasted virtual agent personality assessment across two methods, the standardized Big Five Inventory [32] and an open-ended question in which participants could freely express their first impressions of an agent's personality. This was done for the portrayals of two specific personality traits: Extraversion and Emotional Stability (also known as Neuroticism), which are thought to be more visible than other personality traits. We focused exclusively on the ability of gestures and posture to convey personality,

—————————————————

- *K. Liu is with the Department of Psychology, University of California at Santa Cruz, 1156 High Street, Santa Cruz, CA 95064. Email: kyliu@ucsc.edu*
- *J. Tolins is with the Department of Psychology, University of California at Santa Cruz, 1156 High Street, Santa Cruz, CA 95064. Email: jtolins@ucsc.edu*
- *J. E. Fox Tree is with the Department of Psychology, University of California at Santa Cruz, 1156 High Street, Santa Cruz, CA 95064. Email: foxtree@ucsc.edu*
- *M. Neff is with the Department of Computer Science, 2063 Kemper Hall, University of California, Davis, One Shields Avenue, Davis, CA 95616. Email: mpneff@ucdavis.edu*
- *M. A. Walker is with the Department of Computer Science, Jack Baskin School of Engineering, University of California at Santa Cruz, 1156 High Street, Santa Cruz, CA 95064. Email: maw@soe.ucsc.edu*

without the added confounds of facial expression or linguistic information. Furthermore, we examined scores for all five Big Five traits, as many traits are often intercorrelated (e.g., those who are low on emotional stability may also be lower on agreeableness) [30, 31]. Using the Big Five Inventory (BFI), we confirmed prior researchers' results using the TIPI. We also suggest that using open-ended measures can provide designers with a tool that may offer additional information as to how manipulations of agents' expressive behavior may change user perception of agent personality.

2 PERSONALITY AND PERSONALITY ASSESSMENT IN HUMANS AND VIRTUAL AGENTS

2.1 Assessing Personality

The Big Five model of personality has widespread acceptance in psychology [33, 34], determining human personalities along five dimensions: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism (n.b. Neuroticism is also called Emotional Stability, such that high neuroticism is equivalent to low emotional stability and vice versa). Models of agent behavior have followed suit with assessing personality along the dimensions used for humans [16, 18, 22, 24, 35].

2.2 Open-Ended vs. Forced Choice Question Methods

Human and virtual agent personalities are typically assessed using questionnaires that involve either answers to questions or ratings on behaviors, including adjective rating scales. Popular tools are the Revised NEO Personality Inventory [36], the BFI [32] and TIPI [27]. These have been shown to be reliable for self-evaluation or evaluation by friends or family, but these methods may not accurately capture the first impressions and reactions that are quickly formed by human observers in more naturalistic and situated contexts with strangers [37]. An inventory's exhaustive scope is part of its strength: observers are forced to consider personality facets that might not have occurred to them. But doing so may underestimate the significance of the thoughts that spontaneously arise on their own. If you verbally asked somebody about the personality of a chatty stranger who was blinking more because of dry contact lenses, that person might tell you that the stranger was an extrovert who enjoyed engaging others; he or she may or may not mention anything related to the blinking behavior. However, if you gave him or her a personality inventory that required rating the blinking stranger's emotional stability, the blinking may be retrospectively cast as a nervous tic. In other words, asking specific questions may bias an observer's response, especially because the attribution of personality often happens quickly [1].

This kind of trait or intent attribution has been observed in a variety of communicative domains, both verbal and nonverbal. For example, when asked about what *um* means without priming, people reported that *um*s have something to do with speech production trouble [38]. However, when asked to consider honesty,

people reported that those beginning a turn with *um* were more likely to be lying and less comfortable than those beginning a turn without an *um* [39-41]. Biasing effects have also been found with tone of voice [42].

Similar effects may be present in the work on virtual agents. Participants may be hesitant to ascribe personality to agents [43], or may use the presence of trait-specific questions to infer how they should judge an agent's behavior. This may be particularly relevant in cases where personality tests are trimmed to just those trait questions relevant to a specific manipulation [22, 44].

An additional issue with questionnaire methods is that asking about specific personality traits may under-represent the breadth of personality types observed. In the case of the Big Five, many have argued for other personality dimensions, either replacing existing Big Five dimensions or adding more dimensions currently thought to be overlooked [45-47]. The influence of the questions asked on the elicited personality ratings may be exacerbated for virtual agents, as virtual agents are often less dynamic than humans in movement and expression. An agent may display inconsistent cues for personality, the effects of which may be masked by standard questionnaires [16]. Likert scales allow for *neutral* or *neither agree nor disagree* ratings, but these arguably do not carry the same connotation as *is inconsistent*. Likewise, it is also possible that an observer may think an agent lacks personality. The scales also do not allow additional descriptions such as *robotic* or *creepy*. As virtual agents are likely to be used in contexts in which observers may not necessarily be prone to making in-depth personality judgments, methods for assessing agent personality that tap into initial impressions may provide useful information about the traits that observers find salient enough to comment upon without worrying that they may be biased by trait-specific questions. Open-ended response methods are not a replacement for the thorough personality inventory but rather can convey

Fig. 1. Top: Compact Agent on left, Expansive Agent on right. Bottom: Shaky Agent on left, Smooth Agent on right. Gender was constant within each condition.

complementary data that give those designing virtual agents a fuller idea of how observers view specific agents, particularly during short interactions.

2.3 Contrast Effects

The ability to create multiple identical virtual agents who differ only in highly specific behavioral cues allows for effective assessment of observers' perceptions of personality, using a within-subjects or between-subjects design. Many studies use a within-subjects design where participants are shown multiple agents with different target personality traits, including those that are meant to be the opposites of each other (e.g. extraverted and introverted) [6, 8, 19, 22, 44, 48]. While this method maximizes statistical power and is more logistically efficient, a within-subjects design may introduce unintended contrast effects. Differences found between agents might be driven more by the immediate contrast effects rather than a determination of personality made based on anything an agent was programmed to do. As many agents are ultimately intended to be viewed individually, and not contrastively, assessing personality in the absence of contrast effects using a between-subjects design (i.e., showing only one agent to each participant) would be more in line with the context in which the agent is ultimately used.

Evaluations of personality often rely on the agreement between judgments of either naïve or trained observers. A common methodology within literature exploring the relationship between nonverbal behavior and personality is to have the participants whose nonverbal style is to be judged complete a self-focused personality questionnaire, the results of which are then matched to others' judgments of nonverbal measures [e.g. 49]. The agreement between peers' judgments is typically higher than the agreement between self and peers [33, 50]. The personality that a person thinks they have differs from how others interpret the person's expressive nonverbal behaviors. An analogous mismatch might also exist with virtual agents. Viewers may agree on how to interpret an agent's personality, but that personality may or may not have been intended.

The design of the current study attempts to address some of these possible shortcomings of using a Big Five personality questionnaire for virtual agents, as well as the potential carry-over effects that can accompany a within-subjects design. We asked participants to rate an agent using a standard Big Five personality inventory, but participants were first given an open-ended question, which asked them to qualitatively describe the agent's personality. This taps into the traits that initially seem most salient or noteworthy to the observers and allows them to describe the agent as inconsistently displaying personality or as not having a personality. It also allows participants to mention traits that may not be covered by the Big Five (see [45, 51] for some related discussion of non-Big Five traits). Finally, asking for a qualitative description potentially provides a finer-grained understanding of how agents are being interpreted; for example, an agent who is rated low in emotional stability

TABLE 1
DIFFERENCES IN ANIMATIONS FOR EXPANSIVE AND COMPACT AGENTS

Parameter	Compact Agent	Expansive Agent
Stroke Scale	Slower, smaller, closer to body.	Shorter, larger, further out from body.
Stroke Position Duration	Gestures lower along the vertical axis, closer to Longer gesture durations.	Gestures higher in vertical axis, further from the center Faster gesture durations.
Arm Swivel	N/A	Elbows move away from body during gestures.
Body Motion	Decreased motion in torso and lower body.	N/A
Stance	Shoulders lowered, leaning back.	Shoulders raised, leaning forward.

may be interpreted as angry but not nervous (which both belong to the emotional stability dimension).

Past studies of virtual agent personality judgments have also used a within-subjects design, which maximizes power and experimental efficiency, but does not account for carry-over effects. For instance, a participant may label an agent with narrow gestures as being an introvert because they had previously seen an agent with wide gestures and labeled them an extrovert. However, if they had not seen a purposefully contrasting agent, it is possible that they may have described an agent with narrow gestures as something else entirely, such as *calm* or otherwise high in emotional stability, or perhaps as possessing no standard personality trait. A between-subjects method is costly relative to a within-subjects method, but it does avoid carry-over effects.

2.4 Human Expressions of Extraversion and Emotional Stability

The current study focuses on two traits in the Big Five, extraversion and emotional stability. Of the five traits, these two have been the focus of the most research [52] and have been demonstrated to have reliable visible manifestations. Extroverts make frequent, smooth, and animated movements that are extended away from the body [34, 53, 54]. In comparison, introverts gesture less and their gestures tend to be within a narrower space, with their arms kept close to their bodies. Less emotionally stable people (more neurotic people) make fewer other-directed gestures, more self-directed gestures, more frequent shifts in posture, and lean forward more [55-57]. Another nonverbal bodily pattern observed for some anxious people includes a tense, stiff posture [58]. The more visible a trait, the more accurately it is judged [59]. Because visibility is higher for extraversion, judgments of extraversion are typically more accurate than judgments of neuroticism and the other traits where visibility is lower [33, 59].

2.5 Virtual Agent Expressions of Extraversion and Emotional Stability

Recent work on the expression of extraversion and

emotional stability through gesture in virtual agents has examined the role of gesture and posture in personality attribution by human observers, though they generally tested multiple expressive modalities simultaneously (such as facial expression and gesture [6] or utterance and gesture [18]) rather than examining gesture and posture in isolation.

TABLE 2
DIFFERENCES IN ANIMATIONS FOR SHAKY AND SMOOTH AGENTS

Parameter	Shaky Agent	Smooth Agent
Stroke Scale	Smaller and cross in front of body; jerks added to motion path	Outwards; no jerk added so smooth trajectory.
Stroke Position	Retract position is held out low, to the side of the body.	Hands raised and more in front of the body for relaxed appearance and swing
Arm Swivel Gesture	Elbow rotated inward.	N/A
Phase Connection	Sharper	More rounded
Stance	Collarbones up and further back.	Collarbones down and slightly back.

The 15-second animated clips used in the current study were versions of previously-studied extraverted and introverted agents [22] and agents that are more and less emotionally stable [18] (Figure 3). These earlier studies showed that virtual agents conveyed their intended personality traits when both linguistic and gestural cues were applied. However, a differential perception of emotional stability was only conveyed when self-adaptors (non-communicative, non-interactive touches to one's own body) were included [18]. For this study, we excluded self-adaptors and generated completely new versions of the emotional stability agents that exaggerated the previously used gestures and postures for emotional stability.

In both the extraversion and emotional stability clips, the virtual agents had covered faces to avoid the potentially confounding influence of facial expression. Motion capture data was used for both clips, with edits applied on top to vary qualitative aspects of the motion.

The intended extravert (*Expansive Agent*) contained faster gestures that were larger and located higher and further from the character's centerline. The intended introvert (*Compact Agent*) had slower, narrower gestures that were lower on the torso (Table 1).

For the emotionally stable and unstable clips, the quality of the movement was changed so that the high emotional stability variant (*Smooth Agent*) had a more relaxed stance and produced smoother trajectories while the low emotional stability variant (*Shaky Agent*) had a more tense stance and jerky gestures. The specific manipulations used are summarized in Table 2.

2.6 Current Study

The current study examined whether virtual agents'

gesture and body movement alone conveyed the personalities of extraversion and emotional stability, manipulated separately, and whether the personality perceived varied depending on the method of collecting personality impressions: via surveys (closed, forced-choice questions) or first impressions (open-ended question). The specific open-ended method used in this study provides more nuanced information about an agent's perceived personality that can be used in addition to standard personality questionnaires. Critically, we used a between-subjects paradigm for greater verisimilitude (i.e., agents are often not meant to be used in contrastive pairs so testing them in contrastive pairs may cause unintended carryover effects) and tested whether agents successfully conveyed an intended personality in absence of specific questions that could prime them to see a trait they would not ordinarily see without prompting.

3 EXPERIMENT 1: EXTRAVERSION

Using a between-subjects design, participants saw one of two physically identical agents who were either designed to portray the physical mannerisms of someone who is introverted (Compact Agent) or someone who is extraverted (Expansive Agent). They first described the agent using an open-ended question, and then they assessed the agent's personality using a Big Five inventory.

3.1 Participants

Fifty-nine Amazon Mechanical Turk workers completed both the open-ended question and the personality inventory. Compensation was $0.75 for a task that usually took about 5-7 minutes.

3.2 Procedure

Using a between-subjects design to avoid carry-over effects, the participants watched either a Compact Agent clip or an Expansive Agent clip. They were then immediately asked the open-ended question, *"What personality does this animated character convey to you?"* and given the option of watching the clip over again before answering this question. They were then presented with a modified version of the Big Five Inventory (BFI [32]), a 44-item Big Five assessment that uses 5-point Likert scales. While the original BFI prefaced each item with "*I am someone who…,*" we replaced it with "*If I had to guess, I would describe the character in the video as someone who….*"

3.3 Coding and Scoring

Open-ended responses that only mentioned that the agent appeared unnatural or robotic were counted but were not coded further. Those focused purely on the physical qualities of the agent ("gesturing") or ones that seemed not to be a description of personality ("like a round ball") were also not coded to prevent the blind coders from over-interpreting descriptions that were likely not meant to be about personality. Descriptions that were longer than one word (or phrase) were split into multiple descriptors. Only half of the participants listed two or

TABLE 3
EXAMPLES OF TRAIT CODING BY RATERS

	Condition	Description	Coder 1	Coder 2
Incl.	**Expansive**	Detail-oriented	Organized High Conscientious	Meticulous High Conscientious
		Sociable	Gregarious High Extraversion	Extroverted High Extraversion
	Compact	Chill	Casual High Agreeable	Easygoing High Agreeable
		Geek-like	Aloof Low Extraversion	Seclusive Low Extraversion
Excl.	**Expansive**	Strong	Decisive High Conscientious	Brave High Extraversion
		Responsive	Expressive High Extraversion	Considerate High Agreeable
	Compact	Nonplussed	Curious High Open	Imperceptive Low Conscientious
		Convivial	Jolly High Extraversion	Amiable High Agreeable

Examples of the coders' judgments for descriptions provided by participants that were not already listed by Goldberg (1990). Traits were included if the raters agreed and excluded if they did not.

more descriptors that all fell under a single personality factor; some participants used words that fell under as many as four personality factors. Descriptors were randomized and two blind coders matched them to one of the 339 trait adjectives found in Goldberg's [60] Big Five factor structure. If a description was already listed as a Goldberg adjective loading onto a particular dimension, then the factor it was associated with was counted as the personality factor that the participant found most salient (e.g., "assertive" is associated with high extraversion).

Because few word descriptions matched exact adjectives found in Goldberg's published list, coders chose up to three of the closest synonyms amongst the 339 trait adjectives. To be included in the analysis, adjectives chosen by both coders had to fall under the same personality factor or they were excluded from analysis. For example, the descriptor "arrogant" was kept because one coder chose "boastful" and the other, "pompous" (both low agreeableness). However, for the description "has leadership qualities," one coder chose "dominant" (high extraversion) and another chose "cooperative" (high agreeableness). Such instances of lack of agreement were excluded from the final analysis, as we could not be sure what personality dimension the description was intended to tap into (if they were meant to tap into a conventional Big Five trait at all); thus, no conventional statistical evaluation of inter-rater reliability was undertaken. Table 3 shows examples of what counted as agreement or disagreement in this analysis; descriptors that had agreement were included while those that had disagreement were excluded from analysis.

The BFI was scored according to John, Donahue, and Kentle [32] with personality scores calculated for all five dimensions – openness, conscientiousness, extraversion, agreeableness, and emotional stability (which they call *neuroticism*, i.e. high neuroticism is equivalent to low emotional stability).

3.4 Results

The 59 participants provided 102 personality-related words that were then coded using the Goldberg Adjective List. Nineteen of the 59 (32%) provided single-word answers while six provided descriptions with 4 or more relevant words. The remainder of the participants gave descriptions that counted for 2-3 words. However, of the 102 personality-related words coded, only 57 words (55.9%) were agreed upon by coders to load onto the same personality-related dimension.

Nine of the 59 participants provided descriptions that did not refer to the agent's personality in recognizable terms (e.g. "like a round ball"), including 4 of the 27 Expansive Agent participants (14.8%) and 5 of the 32 Compact Agent participants (15.7%). The agents were equally likely to be described in terms that referred to potential personality traits.

TABLE 4
FREQUENCIES OF TRAIT DESCRIPTIONS FOR EXP 1

	Open		Consci.		Extra.		Agree.		Emo St.	
	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo
Expansive	1	0	1	0	17*	2	5	6	0	0
Compact	1	0	1	0	6	4	11*	2	0	0

Frequencies of Goldberg (1990) adjectives and adjective-equivalents produced by participants by factor. Modal personality trait described is indicated with an asterisk. Open=openness, Consci=conscientiousness, Extra=extraversion, Agree=agreeableness, Emo St.=emotional stability.

3.4.1 Open-Ended Question

For the Expansive Agent, a plurality of participants used descriptions that suggested that they viewed the agent as high in extraversion (17 out of 32 descriptions, or 53.13%). For the Compact Agent, a plurality of participants suggested that they viewed the agent as high in agreeableness (11 out of 25 descriptions, or 44%). Table 4 shows the raw count by personality dimension of the 57 total descriptors that the two coders agreed on.

A Fisher's Exact Test was conducted to see whether there was at least one personality trait that was overwhelmingly associated with a particular type of agent. While there was an obvious plurality of descriptions that favored labeling the Expansive Agent as Highly Extraverted and a slight advantage for labeling the Compact Agent as Highly Agreeable (or, to a lesser extent, Highly Extraverted), the Fisher's Exact Test came close to but did not exceed an alpha of .05 ($p = .054$). However, these results do seem to suggest that there was a slight preference for describing the Expansive Agent in terms related to High Extraversion and the Compact Agent in terms related to High Agreeableness.

3.4.2 Big Five Inventory

BFI ratings of the Compact and Expansive Agents were compared. Because the Big Five Inventory scores for each of the personality factor scales were moderately correlated to each other, we used five separate t-tests, with the Big Five traits as dependent variables and Agent Video as a between-subjects factor. Of the Big Five traits assessed by the BFI, Extraversion was the only trait where participants rated the Compact and Expansive Agents as being different from each other, $t(57) = -4.09$, $p = .0001$, $d = 1.08$, 95% CIs [-1.45, -0.50]. Expansive Agents ($M = 3.61$, $SD = 0.80$) were seen as more extraverted than Compact Agents ($M = 2.64$, $SD = 0.99$). The other trait where participants' ratings potentially differed was Neuroticism, with the Expansive Agent ($M = 3.15$, $SD = 0.86$) being higher in Neuroticism (or lower in Emotional Stability) than the Compact Agent ($M = 2.68$, $SD = 0.81$), $t(57) = -2.13$, $p = .037$, $d = 0.56$, 95% CIs [-0.90, -0.03]; this difference did not reach significance when using the Holm Bonferroni correction for multiple comparisons [29, 30], requiring $p < .0125$. Table 5 shows the mean scores for the five traits.

TABLE 5
BFI RATINGS FOR AGENTS IN EXP 1

	Open	Consci.	Extra.	Agree.	Emo St.
Expansive	3.03 (0.53)	3.23 (0.73)	3.61 (0.80)	2.94 (0.86)	3.15 (0.86)
Compact	2.92 (0.73)	3.18 (0.74)	2.64 (0.99)	3.23 (0.76)	2.68 (0.81)

Means (and SDs) for Big Five ratings by Agent Video.

In short, the Expansive Agents were judged to be more extraverted than the Compact Agents (as they were designed to be). There was a slight (but not statistically significant) tendency for Expansive Agents to also be rated as being less emotionally stable than the Compact Agents, which was unintended.

The boxplot in Figure 2 indicates that when it comes to BFI Extraversion scores, the Expansive Agent did tend to have more tightly clustered results with the narrowest interquartile range, with 50% of scores falling between 3.25 and 4.12 (median = 3.63). That is, a relatively high percentage of participants said at least *somewhat agree* (4) to items on the Extraversion scale.[1] The Compact Agent (median = 2.63), by comparison, with an interquartile range between 2.00 and 3.13, has scores that are more likely to contain *neither agree nor disagree* (3) ratings. While it is safe to say that the Expansive Agent was thought of as at least somewhat extraverted, the statistically significant difference between the Expansive Agent's and the Compact Agent's Extraversion scores does not necessarily mean that the Compact Agent was thought of as introverted. Instead, the Compact Agent may have been thought of as neutral.

3.5 Discussion

As measured by the BFI, the Expansive Agent was more likely to convey extraversion than the Compact Agent (as intended). While we can somewhat safely conclude that the Expansive Agent was seen as extraverted, it is less clear-cut whether the Compact Agent was seen as introverted. A rating of 3 is a neutral rating and neither agent was overwhelmingly rated as being strongly extraverted or introverted. Instead, the Expansive Agent got higher ratings on the BFI Extraversion scale than the Compact Agent, whose mean scores were close to a neutral rating. As a result, it may be said that Expansive Agent was extraverted, but that does not make the Compact Agent introverted.

The open-ended question provided a somewhat complementary picture. The Expansive Agent was more frequently described as being highly extraverted. The Expansive Agent was also consistently ascribed extraversion-related descriptions over other personality trait descriptions. The Compact Agent was far less frequently described in terms of extraversion. Instead, participants viewing the Compact Agent tended to use agreeableness descriptions.

There was agreement across the BFI and open-ended responses that the Expansive Agent was seen as being more extraverted than the Compact Agent. However, neither of the agents were seen as strikingly extraverted or introverted, even though extraversion is thought to be a personality trait that is more likely than most others to manifest in gesture. While there was a small difference in the BFI ratings for Neuroticism between the two agents, few emotional stability-related descriptions were used in the open-ended response. In contrast, although Agreeableness was frequently mentioned in the open-ended descriptions, participants tended to rate both agents similarly in terms of Agreeableness on the BFI, suggesting that it is a trait that jumps out at them initially but does not necessarily last. This could be partially due to the time elapsed between watching the video and completing the 44-item BFI: participants were allowed to watch it multiple times in the beginning but could not
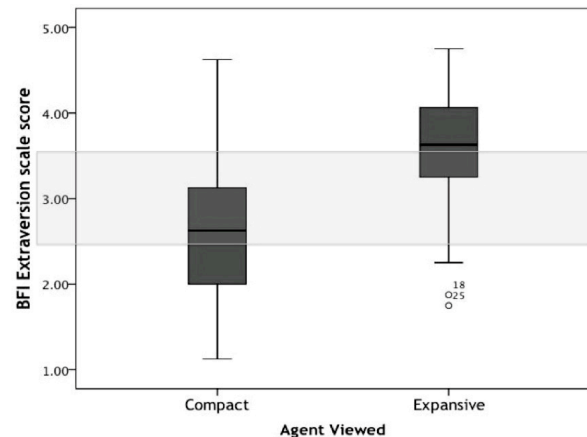

Fig. 2. Boxplot for the distribution of Extraversion scores for the two agents. The grey area denotes range where many responses were likely to be *neither agree nor disagree*.

---

[1] Of the eight items that comprise the BFI Extraversion scale, only half showed a significant effect of the video clip viewed using a Bonferroni-corrected $p = .006$: Reserved, Full of Energy, Generates Enthusiasm, and Assertive.

watch it again after they moved to answering questions. The nature of the Agreeableness trait may also play a role in the seemingly transient reaction. Agreeableness is sometimes called *friendliness* or *likability* [36, 61] and it is possible that participants are more likely to integrate judgments of personal liking with personality judgments when providing an open-ended response than in the BFI. Previous work on humans indeed shows that first impressions for some traits are more likely to change than others [4]. The BFI Agreeability scale includes items such as *Is considerate and kind to almost everyone*. A 15 second clip may not be enough time to convey this level of information, but it may be sufficient time to evoke a less specific description, such as "She seems friendly."

TABLE 6
FREQUENCIES OF TRAIT DESCRIPTIONS FOR EXP 2

	Open		Consci.		Extra.		Agree.		Emo St.	
	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo
Smooth	4	2	1	1	14*	1	5	4	1	1
Shaky	1	1	0	5	8	6	7	17*	0	5

Frequencies of Goldberg (1990) adjectives and adjective-equivalents produced by participants by factor. Modal personality trait described is indicated with an asterisk.

4 EXPERIMENT 2: EMOTIONAL STABILITY

Using a between-subjects design, participants saw one of two physically identical agents who were either designed to portray the physical mannerisms of someone who is high in emotional stability (Smooth Agent) or someone who is low in emotional stability (Shaky Agent). They first described the agent using an open-ended question, and then they assessed the agent's personality using a Big Five inventory.

4.1 Participants

There were a total of 84 participants in all (Shaky N = 49, Smooth N = 35): 26 (30.95%) were recruited through UCSC's participant pool for class credit, 52 (65.82%) were Mechanical Turk workers awarded $1, and the remainder was collected by convenience sampling of research assistants and their friends who were blind to the experiment. All 84 participants completed the open-ended question and the BFI inventory.

4.2 Procedure

The procedure was the same as Experiment 1, except that participants viewed stimuli meant to convey the opposing poles of emotional stability, rather than extraversion.

4.3 Coding and Scoring

The procedure was the same as Experiment 1.

4.4 Results

4.4.1 Open-Ended Question

Although all participants were asked to provide a description of the agent's personality using the open-ended question, not all provided usable, personality-related descriptions. Fifteen (of 49, or 30.61%) Shaky Agent participants and 7 (of 35, or 20%) Smooth Agent participants did not provide a usable response: 8 Shaky Agent participants and 4 Smooth Agent participants either did not answer or provided a response that had nothing to do with the agent (for example, responses related to video quality or Sonic the Hedgehog). An additional 7 Shaky Agent participants and 3 Smooth Agent participants did not provide descriptions related to personality (for example, descriptions such as "character is gesturing").

The remaining 62 participants (N = 28 for Smooth Agents and N = 34 for Shaky Agents) provided 121 personality-related words that were then coded using Goldberg Adjective List. Only 3 participants provided descriptions that used 3 or more personality-relevant words, while 19 participants gave single-word responses. Thirty-seven descriptors were dropped from the analysis because disagreements in coding suggested that they were either not Big Five personality traits or were too ambiguous to fit into a single Goldberg [60] factor cluster. Table 6 shows the frequencies of traits described by participants.

TABLE 7
BFI RATINGS FOR AGENTS IN EXP 2

	Open	Consci.	Extra.	Agree.	Emo St.
Shaky	3.11 (0.95)	2.65 (0.92)	3.15 (0.72)	3.39 (0.95)	2.91 (0.78)
Smooth	3.39 (0.85)	3.16 (0.85)	3.46 (0.59)	2.76 (0.79)	3.09 (0.63)

Mean (and standard deviation) BFI scores for agents by Agent Video

The Shaky and Smooth Agents elicited descriptions across all five personality dimensions, though descriptions also seemed to focus on a couple of specific traits. A Fisher's Exact Test revealed that participants' spontaneous descriptions of a personality did differ depending on whether participants watched either the Smooth or the Shaky Agent (p = .01). The most frequent personality trait ascribed by those observing the Smooth Agent was high extraversion. The most frequent personality trait ascribed by those observing the Shaky Agent was low agreeableness. Emotional stability was infrequently mentioned, though the Shaky Agent was labeled as being low in emotional stability more often than the Smooth Agent, who was rarely talked about in those terms.

Participants seemed to have not spontaneously labeled the Smooth Agent as being emotionally stable, or did not find it noteworthy enough to note. Instead, they described the Smooth Agent as an extraverted personality. On the other hand, the Shaky Agent was read by some as conveying an emotionally unstable personality, but most adjectives described the agent as

being low in agreeableness; that is, emotional stability was not the most salient trait commented on.

4.4.2 Big Five Inventory

An independent-samples t-tests were run using the Big Five traits as dependent variables and Agent Viewed (Smooth/Shaky) as the between-subjects factor. Of these five, two traits showed a significant difference between Smooth and Shaky Agents when using a Holm-Bonferroni corrected critical p-value. The difference in Conscientiousness between Shaky and Smooth Agents trended towards significant. Specifically, the Shaky Agent ($M = 3.39$, $SD = 0.95$) was rated as higher on Neuroticism (or lower in emotional stability) than Smooth Agent ($M = 2.76$, $SD = 0.79$), $t(82) = 3.16$, $p = .002$, $d = 0.72$, 95% CIs [0.23, 1.01]. The Shaky Agent ($M = 2.65$, $SD = 0.92$) was also seen as being lower on Agreeableness than the Smooth Agent ($M = 3.16$, $SD = 0.85$), $t(82) = -2.57$, $p = .01$, $d = .58$, 95% CIs [-0.90, -0.12]. Finally, the Shaky Agent ($M = 3.15$, $SD = 0.72$) was seen as being lower in Conscientiousness than the Smooth Agent ($M = 3.46$, $SD = 0.59$), $t(82) = -2.08$, $p = .04$, $d = .47$, 95% CIs [-0.60, -0.01], but this did not exceed the corrected critical value of $p < .017$.

This suggests that, as intended, the Shaky Agent was judged as less emotionally stable than the Smooth Agent. (Incidentally, this also indicates that the exaggeration of the gesture edits from those used in Neff et al. [18] was effective, as the previous study did not find a difference in emotional stability from gesture variation alone.) At the same time, the Shaky Agent's behaviors were also read as being less agreeable than the Smooth Agent's, which was unintended. There was also a slight but non-significant difference in conscientiousness. Table 7 shows BFI ratings for all five traits.
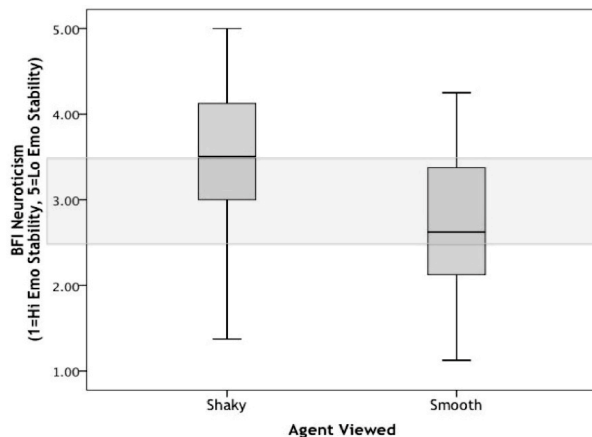
Fig. 3. Boxplot for the distribution of BFI Neuroticism ratings (Emotional Stability) by Agent Video. The grey area denotes range where many responses were likely to be *neither agree nor disagree*.

The boxplot shown in Figure 3 shows the distribution of BFI Neuroticism scores (1 is high emotional stability and 5 is low emotional stability) for the Shaky and Smooth Agents. Though there is a narrower distribution for the Smooth Agents, with few participants *strongly*

*agree*ing that the Smooth Agent was low in emotional stability, the bars are still quite wide for both agents.[2] Though there is a difference in medians between Smooth and Shaky (2.63 vs. 3.50, respectively) and interquartile ranges' overlap (2.13 - 3.50 vs. 2.94 - 4.13) suggests that around a quarter of participants did not distinguish between the two agents on this trait; their scores are likely to be closer to *neutral* than otherwise. Nevertheless, there was a preference for seeing the Shaky Agent as being low in emotional stability and the Smooth Agent as being high in emotional stability.
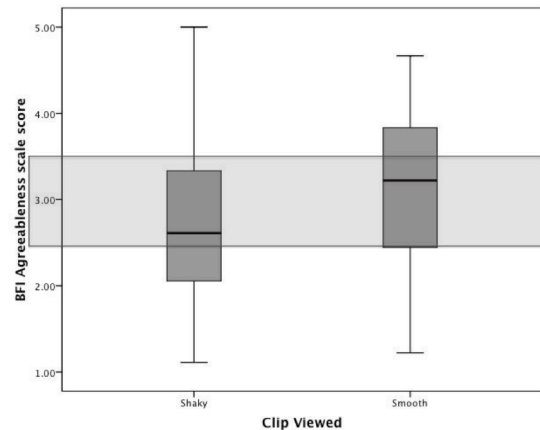
Fig. 4. Boxplot for the distribution of BFI Agreeableness by Agent Video. Again, the grey area denotes range where many responses were likely to be *neither agree nor disagree*.

Figure 4 shows the boxplots for Agreeableness, which indicate that even though neither agent was rated as particularly agreeable or disagreeable (nor were their ratings that different from each other), participants tended to *somewhat disagree* with Agreeableness items for the Shaky Agent. The median (3.22) and interquartile range (2.00-3.89) for the Smooth Agent suggest that the agent cannot be consistently characterized as either low or high in agreeableness, even if there were a statistically significant difference between the means.

4.4 Discussion

The open-ended responses showed that the most frequently ascribed personality trait for the Shaky Agent was low agreeableness while the most frequently ascribed trait for the Smooth Agent was high extraversion; neither agent was very commonly described in terms of emotional stability (though it was slightly more common for the Shaky Agent than the Smooth Agent). It is interesting to note that in previous work [18], the presence of self-adaptors was the movement factor that produced a significant difference between emotional stability ratings, with self-adaptors leading people to rate the agent as more neurotic. That movement cue was not

---

[2] Of the eight items that comprise the BFI Neuroticism scale, only two showed significant effect of the video clip viewed using a Bonferroni-corrected $p = .006$: Emotionally Stable and Can Be Moody.

included in this work, but may have a strong influence and could have led to different open-ended responses. The BFI agreeableness scores somewhat supported the open-ended responses, indicating that people were more likely to see the Shaky Agent as being lower in agreeableness than the Smooth Agent. Yet while neither agent was qualitatively described as being particularly high or low in emotional stability, the BFI Neuroticism score indicated that there was a difference between the Shaky and Smooth Agents: namely, the Shaky Agent was seen as being lower in emotional stability than the Smooth Agent. Participants seemed a little more convinced of the Shaky Agent's lack of emotional stability than of the Smooth Agent's abundance of it.

Interestingly, the Smooth Agent was described in terms relating to high extraversion in the open-ended responses but was not rated as being extraverted on the BFI. It seems unlikely that, absent any other salient trait, participants were defaulting to an assumption that extraverted people gesture more than introverted ones, as descriptions relating to high extraversion were not generally found in the Experiment 1 open-ended responses (even though the agents were intended to convey an introvert and an extravert). The descriptions that suggest that the Smooth Agent was generally higher in agreeableness was supported by the mean differences between Shaky and Smooth Agents. This, however, is tempered by the fact that when asked in the BFI, participants generally responded with neutral ratings on Agreeableness items, suggesting that participants did not view them as strongly agreeable or not agreeable.

5 GENERAL DISCUSSION

Nonverbal bodily behaviors are one of numerous channels with which people express and perceive personality. Although personality is conveyed through words, prosody, facial expressions and other behaviors in addition to nonverbal bodily expressions, impressions are still easily – though less accurately – formed when given only one of these channels [62].

It is critical for the successful implementation of agents across applications that they be perceived as coherent individuals, such that their behavior may be interpreted using the same social strategies employed in social interaction with humans [12]. As such, designers need both models of how personalities are expressed across various modalities, which may be borrowed wholesale from the psychological literature, and also measures to test whether observers indeed perceive the intended personalities. We analyzed two such methods for exploring how manipulations of expressive nonverbal style lead to changes in the judgment of personality: questionnaire-style personality inventory and an open-ended question. We used a between-subjects technique that purposefully avoided contrast and carry-over effects. Many agents are presumably not intended for use as contrastive pairs of opposite personalities, making it necessary to see if virtual agent personality can be conveyed when each agent is viewed in isolation.

Similar to the current study, previous work on the attribution of Big Five personality dimensions to virtual agents has also shown a relationship between the attribution of emotional stability and the attribution of agreeableness by observers to agents under certain conditions [18]. Other previous studies, however, focused exclusively on a single target personality dimension, to the point of reducing personality surveys to just those scales measuring that dimension [22, 43, 44]. This confirmatory bias is in conflict with the generally accepted conceptualization of personality as a unified whole across the traits. This study, as well as others (e.g. [6]), demonstrate that manipulation along a single dimension may not be perceived as such by users. Understanding how these unintended perceptions are produced and whether they are reliable should be a focus of future work. In the creation and control of task- or user-specific agents, these effects are critically important.

Virtual agents can lend themselves to some descriptions that might be less likely to be ascribed to humans, such as *robotic*. Yet there is reason to believe that methods used with humans, such as the BFI, are appropriate for the assessment of virtual agent personality. People not only ascribe personality traits to other people based on very little information, they ascribe human behaviors to inanimate things. We can observe this facility even in the least human of stimuli such as Heider and Simmel's silent, black-and-white shapes "chasing" each other [63]. Although a personality-related description was not always spontaneously offered, these shapes were ascribed the intentionality of animate, if not human, objects. In our work with virtual agents, we found that many of the responses often mentioned some sort of goal-directed, intentional behavior, even if there was no description of personality or distinct personality traits per se: the agents were not infrequently described as *explaining*, *asking*, or *instructing.*

In two experiments, we showed that the standardized personality inventory and open-ended question methods of assessing personality are complementary. Open-ended questions allow participants to express their perceptions without priming them on any particular dimension of the Big Five. This method is useful for obtaining a sense of what traits the participants thought were the most relevant and visible. It reflects a naturalistic process for the ascription of personality, which is far more free-form and unstructured than the structured querying of an inventory. Personality inventories, on the other hand, necessarily allow probing of multiple personality dimensions, which can be an advantage when observers forget to mention something they may have genuinely observed. Thoughtful choice and interpretation of the method (or methods) of assessment are necessary, as different methods can provide very different information.

Together, the inventory and open-ended methods provide evidence that opposite gestures are not equivalent to opposite poles of personality dimensions. For example, both open-ended responses and BFI ratings demonstrated that performing the opposite of a certain behavior does not necessarily lead to an equivalent

judgment of an opposing personality trait. The agent that used wide gestures (Expansive) was seen as at least somewhat extraverted, but the agent that used narrow gestures (Compact) was not seen as being particularly introverted.

Each method, when used in isolation, had advantages and disadvantages. While more labor-intensive, an open-ended question allows a more detailed picture of an agent's personality to emerge. Although the individual items of a Big Five scale can be examined on a longer assessment such as the 44-item BFI, this may not always be that informative. For example, in Experiment 2, the differences between the Smooth and Shaky Agents' emotional stability scores were driven by ratings on two items: *Is emotionally stable, not easily upset* and *Can be moody*. These are somewhat vague (and sometimes double-barreled) descriptions, yet people would probably approach someone with a short fuse differently than they would a worrier. That is, with just the BFI, we don't know which of the more precise behaviors dominated an observer's interpretation. With the open-ended question, we do know. A few participants described the Shaky Agent as *anxious* or *nervous*, but more described the agent as *mean*, *angry*, *short-tempered*, and *aggravated*. We also know more than the BFI allows. Agents were commonly described as *confused*, for example, which was not consistently coded into any Big Five category. In other words, qualitative data and closer scrutiny of score distribution and a scale's individual items (when possible) can help flesh out numbers that might otherwise gloss over important distinctions or even tap into alternate interpretations of intended manipulations.

Open-ended measures are helpful when an agent's personality ratings on an inventory largely fall into "neutral" territory, which is what we saw in Experiment 2, where the Shaky Agent was seen as less emotionally stable, on average, than the Smooth Agent (which replicates previous findings by Neff and colleagues [18, 22]) but the spread of the Smooth Agent's ratings were mostly contained within *neither agree nor disagree* on the BFI's Neuroticism items. While researchers can (and have) shown that there are mean differences in personality ratings when observers view multiple contrastive agents, it is also useful to examine open-ended descriptions of agents viewed in isolation. This gives those who design agents a clearer idea of how their agents are seen in the event that their agents are not yet hitting intended personality notes hard enough to register consistently on standardized inventories.

More work, however, will need to be done to understand why participants chose to mention certain personality dimensions, particularly when those dimensions are not thought to be particularly visible in gesture (such as agreeableness): Was this because our participants were used to talking about whether others are friendly or unfriendly, because the gestures we used really tapped into actual gestural correlates of agreeableness, or because another trait such as emotional stability mediated judgments of agreeableness (i.e., those who come off as angry were also thought to be less friendly)? There are also cultural considerations that are outside the scope of this current study that should be investigated in future work. Other work shows that the relatively new ability of TTS engines to produce voices with dialectical variations leads to stereotypical attributions to agents with those voices: for example, an agent that speaks with a regional accent has a sense of humor, whereas an agent that speaks with Received Pronunciation is more educated and trustworthy [11].

Future work should also involve the direct comparison of the perception of personality traits expressed by agents of different genders as the current study did not examine gender differences. Previous work for example has found task-specific effects for gender, where female agent voices are viewed as more competent at prototypically female tasks [14]. There may also be additional physical cues, such as clothing or physique, that could influence personality perception when gestures are added [64]. Finally, though we find it likely that additional contextual, physical, and expressive cues will change interpretations of agent behavior, as they change interpretations of human behavior, our study can make no claim regarding how additional factors would influence personality perception.

The Big Five and its attendant personality inventories are, arguably, most powerful when putting individuals into analyzable categories in order to describe the average behavior of a group [34]. Nevertheless, such inventories may be less effective when trying to create an agent that expresses a particular personality, as their scales and individual items are designed with wording that is intentionally terse and efficient. Open-ended descriptions do not have to be highly consistent to offer guidance on how to design agents in a future iteration to more strongly convey a target personality trait. Even when agents differed on many items on a given scale, as the Expansive and Compact Agents did on Extraversion, it can still be useful to know that Expansive Agent was seen by some as *excited* and *outgoing*, though also *bossy*, *angry*, and never *friendly*. The Compact Agent elicited many descriptions like *calm*, *laid-back*, *nonchalant*, *relaxed,* and *easy-going*. Though a more laborious process, this qualitative type of information can help in the creation of agents with a cohesive suite of expressive and stylistic behaviors, based on user perception of prescribed personality parameters. Accurately applying such personality models and testing them with appropriate measures allows for the successful implementation of expressive agents with cohesive, believable personalities.

6 CITATIONS AND ACKNOWLEDGMENTS

6.1 Acknowledgments

6.2 Citations

1. N. Ambady, M. Hallahan and R. Rosenthal, "On judging and being judged accurately in zero-acquaintance situations," *Journal of Personality and Social Psychology*, vol. 69, no. 3, 1995, pp. 518.

2. M.J. Levesque and D.A. Kenny, "Accuracy of behavioral predictions at zero acquaintance: A social relations analysis," *Journal of Personality and Social Psychology*, vol. 65, no. 6, 1993, pp. 1178.

3. L. Ross, "The intuitive psychologist and his shortcomings: Distortions in the attribution process," *Advances in experimental social psychology*, vol. 10, 1977, pp. 173-220.

4. L.K. Kammrath, D.R. Ames and A.A. Scholer, "Keeping up impressions: Inferential rules for impression change across the big five," *Journal of Experimental Social Psychology*, vol. 43, no. 3, 2007, pp. 450-457.

5. E. Campana, M.K. Tanenhaus, J.F. Allen and R.W. Remington, "Evaluating cognitive load in spoken language interfaces using a dual-task paradigm," *Proc. INTERSPEECH*, 2004.

6. M. McRorie, I. Sneddon, G. McKeown, E. Bevacqua, E. de Sevin and C. Pelachaud, "Evaluation of four designed virtual agent personalities," *Affective Computing, IEEE Transactions on*, vol. 3, no. 3, 2012, pp. 311-322.

7. A. Tapus and M.J. Mataric, "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance," *Proc. AAAI Spring Symposium: Emotion, Personality, and Social Behavior*, 2008, pp. 133-140.

8. E. Bevacqua, E. de Sevin, C. Pelachaud, M. McRorie and I. Sneddon, "Building credible agents: Behaviour influenced by personality and emotional traits," *Proc. Kansei Engineering and Emotion Research, Paris, France*, 2010.

9. I. Damian, T. Baur and E. André, "Investigating social cue-based interaction in digital learning games," *Proc. Proceedings of the 8th International Conference on the Foundations of Digital Games, SASDG*, 2013.

10. I. Damian, B. Endrass, P. Huber, N. Bee and E. André, "Individualized agent interactions," *Motion in games*, Springer, 2011, pp. 15-26.

11. B. Krenn, S. Schreitter, F. Neubarth and G. Sieber, "Social evaluation of artificial agents by language varieties," *Proc. Intelligent Virtual Agents*, Springer, 2012, pp. 377-389.

12. B. Reeves and C. Nass, "The media equation: How people treat computers, television,? New media like real people? Places," *Computers & Mathematics with Applications*, vol. 33, no. 5, 1997, pp. 128-128.

13. F. Mairesse and M.A. Walker, "Towards personality-based user adaptation: Psychologically informed stylistic language generation," *User Modeling and User-Adapted Interaction*, vol. 20, no. 3, 2010, pp. 227-278.

14. D. Kuchenbrandt, M. Häring, J. Eichberg, F. Eyssel and E. André, "Keep an eye on the task! How gender typicality of tasks influence human–robot interactions," *International Journal of Social Robotics*, vol. 6, no. 3, 2014, pp. 417-427.

15. K. Isbister and C. Nass, "Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics," *International journal of human-computer studies*, vol. 53, no. 2, 2000, pp. 251-267.

16. F. Mairesse and M.A. Walker, "Controlling user perceptions of linguistic style: Trainable generation of personality traits," *Computational Linguistics*, vol. 37, no. 3, 2011, pp. 455-488.

17. A.J. Gill, C. Brockmann and J. Oberlander, "Perceptions of alignment and personality in generated dialogue," *Proc. Proceedings of the Seventh International Natural Language Generation Conference*, Association for Computational Linguistics, 2012, pp. 40-48.

18. M. Neff, N. Toothman, R. Bowmani, J.E. Fox Tree and M.A. Walker, "Don't scratch! Self-adaptors reflect emotional stability," *Proc. Intelligent Virtual Agents*, Springer, 2011, pp. 398-411.

19. N. Bee, C. Pollock, E. André and M. Walker, "Bossy or wimpy: Expressing social dominance by combining gaze and linguistic behaviors," *Proc. Intelligent Virtual Agents*, Springer, 2010, pp. 265-271.

20. S. Kopp, et al., "Towards a common framework for multimodal generation: The behavior markup language," *Proc. Intelligent virtual agents*, Springer, 2006, pp. 205-217.

21. A. Egges, S. Kshirsagar and N. Magnenat-Thalmann, "Generic personality and emotion simulation for conversational agents," *Computer Animation and Virtual Worlds*, vol. 15, no. 1, 2004, pp. 1-13.

22. M. Neff, Y. Wang, R. Abbott and M. Walker, "Evaluating the effect of gesture and language on personality perception in conversational agents," *Proc. Intelligent Virtual Agents*, Springer, 2010, pp. 222-235.

23. C. Pelachaud, "Studies on gesture expressivity for a virtual agent," *Speech Communication*, vol. 51, no. 7, 2009, pp. 630-639.

24. F. Mairesse, M.A. Walker, M.R. Mehl and R.K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *J. Artif. Intell. Res.(JAIR)*, vol. 30, 2007, pp. 457-500.

25. A. Cafaro, H.H. Vilhjálmsson, T. Bickmore, D. Heylen, K.R. Jóhannsdóttir and G.S. Valgarðsson, "First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters," *Proc. Intelligent Virtual Agents*, Springer, 2012, pp. 67-80.

26. C. Faur, C. Clavel, S. Pesty and J.-C. Martin, "Perseed: A self-based model of personality for virtual agents inspired by socio-cognitive theories," *Proc. Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, IEEE, 2013, pp. 467-472.

27. S.D. Gosling, P.J. Rentfrow and W.B. Swann, "A very brief measure of the big-five personality domains," *Journal of Research in personality*, vol. 37, no. 6, 2003, pp. 504-528.

28. G. Saucier, "Mini-markers: A brief version of goldberg's unipolar big-five markers," *Journal of personality assessment*, vol. 63, no. 3, 1994, pp. 506-516.

29. Z. Callejas, D. Griol and R. López-Cózar, "A framework for the assessment of synthetic personalities according to user perception," *International Journal of Human-Computer Studies*, vol. 72, no. 7, 2014, pp. 567-583.

30. A. Panter, J. Tanaka, R. Hoyle, C. Halverson, G. Kohnstamm and R. Martin, "Structural models for multimode designs in personality and temperament research," *The developing structure of temperament and personality from infancy to adulthood*, 1994, pp. 111-138.

31. J.M. Digman, "The curious history of the five-factor model," 1996.

32. O.P. John, E.M. Donahue and R.L. Kentle, "The big five inventory—versions 4a and 54," *Berkeley: University of California, Berkeley, Institute of Personality and Social Research*, 1991.

33. O.P. John and R.W. Robins, "Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self," *Journal of Personality*, vol. 61, no. 4, 1993, pp. 521-551.

34. B.H. LaFrance, A.D. Heisel and M.J. Beatty, "Is there empirical evidence for a nonverbal profile of extraversion?: A meta-analysis and critique of the literature," *Communication Monographs*, vol. 71, no. 38-48, 2004.

35. J. Allbeck and N. Badler, "Toward representing agent behaviors modified by personality and emotion," *Embodied Conversational Agents at AAMAS*, vol. 2, 2002, pp. 15-19.

36. P.T. Costa and R.R. McCrae, ""Normal'personality inventories in clinical assessment: General requirements and the potential for using the neo personality inventory": Reply," 1992.

37. L. Albright, A.I. Cohen, T.E. Malloy, T. Christ and G. Bromgard, "Judgments of communicative intent in conversation," *Journal of Experimental Social Psychology*, vol. 40, 2004, pp. 290–302.

38. J.E. Fox Tree, "Folk notions of um and uh, you know and like.," *Text & Talk*, vol. 27, no. 3, 2007, pp. 297-314.

39. J.E. Fox Tree, "Interpreting pauses and ums at turn exchanges," *Discourse Processes*, vol. 34, no. 1, 2002, pp. 37-55.

40. L.A. Hosman, T.M. Huebner and S.A. Siltanen, "The impact of power-of-speech style, argument strength, and need for cognition on impression formation, cognitive responses, and persuasion " *Journal of Language and Social Psychology*, vol. 21, 2002.

41. H.H. Clark and J.E. Fox Tree, "Using uh and um in spontaneous speaking," *Cognition*, vol. 84, no. 1, 2002, pp. 73-111.

42. G.A. Bryant and J.E. Fox Tree, "Recognizing verbal irony in spontaneous speech," *Metaphor and symbol*, vol. 17, no. 2, 2002, pp. 99-119.

43. K. Van den Bosch, A. Brandenburgh, T.J. Muller and A. Heuvelink, "Characters with personality!," *Proc. Intelligent Virtual Agents*, Springer, 2012, pp. 426-439.

44. D. Arellano, J. Varona, F.J. Perales, N. Bee, K. Janowski and E. André, "Influence of head orientation in perception of personality traits in virtual agents," *Proc. The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, International Foundation for Autonomous Agents and Multiagent Systems, 2011, pp. 1093-1094.

45. S.V. Paunonen and D.N. Jackson, "What is beyond the big five? Plenty!," *Journal of personality*, vol. 68, no. 5, 2000, pp. 821-835.

46. M.C. Ashton, et al., "A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages," *Journal of Personality and Social Psychology*, vol. 86, no. 2, 2004, pp. 356.

47. M. Zuckerman, D.M. Kuhlman, J. Joireman, P. Teta and M. Kraft, "A comparison of three structural models for personality: The big three, the big five, and the alternative five," *Journal of Personality and Social Psychology*, vol. 65, no. 4, 1993, pp. 757.

48. J. Kim, S.S. Kwak and M. Kim, "Entertainment robot personality design based on basic factors of motions: A case study with rolly," *Proc. Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, IEEE, 2009, pp. 803-808.

49. P.E. Gallaher, "Individual differences in nonverbal behavior: Dimensions of style," *Journal of Personality and Social Psychology*, vol. 63, no. 1, 1992, pp. 133.

50. M.R. Mehl, J.W. Pennebaker, D.M. Crow, J. Dabbs and J.H. Price, "The electronically activated recorder (ear): A device for sampling naturalistic daily activities and conversations," *Behavior Research Methods, Instruments, & Computers*, vol. 33, no. 4, 2001, pp. 517-523.

51. K. Lee, B. Ogunfowora and M.C. Ashton, "Personality traits beyond the big five: Are they within the hexaco space?," *Journal of personality*, vol. 73, no. 5, 2005, pp. 1437-1463.

52. R.M. Tobin, W.G. Graziano, E.J. Vanman and L.G. Tassinary, "Personality, emotional experience, and efforts to control emotions," *Journal of Personality and Social Psychology*, vol. 79, no. 4, 2000, pp. 656.

53. R.E. Riggio and H.S. Friedman, "Impression formation: The role of expressive behavior," *Journal of Personality and Social Psychology*, vol. 50, no. 2, 1986, pp. 421-427.

54. R. Lippa, "The nonverbal display and judgment of extraversion, masculinity, femininity, and gender diagnosticity: A lens model analysis," *Journal of Research in Personality*, vol. 32, no. 1, 1998, pp. 80-107.

55. M. Argyle, *Bodily communication*, Routledge, 2013.

56. A. Campbell and J.P. Rushton, "Bodily communication and personality," *British Journal of Social and Clinical Psychology*, vol. 17, no. 1, 1978, pp. 31-36.

57. P. Feyereisen and J.-D. de Lannoy, *Gestures and speech: Psychological investigations*, Cambridge University Press, 1991.

58. P.H. Waxer, "Nonverbal cues for anxiety: An examination of emotional leakage," *Journal of abnormal psychology*, vol. 86, no. 3, 1977, pp. 306.

59. D.C. Funder and K.M. Dobroth, "Differences between traits: Properties associated with interjudge agreement," *Journal of Personality and Social Psychology*, vol. 52, no. 2, 1987, pp. 409-418.

60. L.R. Goldberg, "An alternative "description of personality": The big-five factor structure," *Journal of Personality and Social Psychology*, vol. 59, no. 6, 1990, pp. 1216-1229.

61. J.M. Digman, "Personality structure: Emergence of the five-factor model," *Annual review of psychology*, vol. 41, no. 1, 1990, pp. 417-440.

62. P. Ekman, W.V. Friesen, M. O'Sullivan and K. Scherer, "Relative importance of face, body, and speech in judgments of personality and affect," *Journal of Personality and Social Psychology*, vol. 38, no. 2, 1980, pp. 270.

63. F. Heider and M. Simmel, "An experimental study of apparent behavior," *The American Journal of Psychology*, 1944, pp. 243-259.

64. L.P. Naumann, S. Vazire, P.J. Rentfrow and S.D. Gosling, "Personality judgments based on physical appearance," *Personality and Social Psychology Bulletin*, 2009.

65. K. Liu, J. Tolins, J.E. Fox Tree, M. Walker and M. Neff, "Judging iva personality using an open-ended question," *Proc. Intelligent Virtual Agents*, Springer, 2013, pp. 396-405.