

Judging IVA Personality Using an Open-Ended Question

Kris Liu¹, Jackson Tolins¹, Jean E. Fox Tree¹, Marilyn Walker¹, Michael Neff²

¹ University of California, Santa Cruz

{kylui|jtolins|foxtree|maw}@ucsc.edu, maw@soe.ucsc.edu

² University of California, Davis

mpneff@ucdavis.edu

Abstract. Judgments of personality typically employ ratings of Big Five scale items such as “How emotionally stable is this person?” with choices from 1 (least) to 5 (most). Such questions focus raters’ attention on an experimenter’s dimensions of interest. We show that the personality traits provided as a result of open-ended questions such as “What personality does this animated character convey to you?” can differ from those observed when raters are given scale questions. Using IVAs that gesture in ways associated with emotionally stable and unstable people, we showed that participants were more likely to describe the unstable agent as disagreeable and the stable agent as extraverted; emotional stability was not usually mentioned. However, a Big Five inventory showed that these agents differed on agreeableness and emotional stability. An open-ended question method of assessing what personality an IVA conveys can potentially be more informative than using scale-item inventories alone.

Keywords: personality, Big Five, methodology, gesture, non-verbal behavior

1 Introduction

Measuring an individual’s personality typically relies on the descriptive adjectives used to develop the Big Five or OCEAN model [13, 16, 19], which claims that personality varies on five factors: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism (which we refer to as emotional stability, such that high neuroticism is low emotional stability and vice versa). The Big Five model of personality has become a standard in the field of psychology over the last 50 years. A large body of research has focused on the creation of personality questionnaires, inventories, and adjective rating scales designed to measure these broad dimensions. The most well used personality inventory to develop from this methodology is the Revised NEO Personality Inventory (NEO-PI-R), developed by Costa and McCrae [4]. But there have been other valid and successful measures such as the Big Five Inventory (BFI), developed by John, Donahue, and Kentle [8], and Goldberg’s trait descriptive adjective [7] lists, which has high internal consistency and has been easily replicated [10].

Two related problems arise when using any of a number of Big Five inventories for the purpose of asking an observer to assess the behavior of another. First, the characterizations of personality that are spontaneously generated by people may be different from measurements gleaned from personality inventories, as people may be primed or influenced by the wording of scale rating items such as “This person is emotionally stable.” Second, behaviors not associated with anything having to do with personality may appear diagnostic when considered in light of a personality trait. For example, touching one’s hair is not always a sign of anxiety. But if an observer is asked whether one person who is touching his hair is more anxious than another person who is not, that behavior may now seem indicative of anxiety even if the observer did not previously mention hair-touching as something that anxious people do.

A linguistic analogue can be found in *um*, which is often believed to indicate anxiety [2, 3]. Listeners judge speakers who use *ums* with a particular topic as less comfortable than those who do not [5]. Speakers are also seen as more dishonest and more likely to be experiencing speech production problems. The fact that listeners are willing to judge those utterances as indicating more anxiety, dishonesty and production trouble than non-*um* utterances suggests not only that there are multiple impressions an *um* can convey, but that listeners are willing to make various judgments depending on the question asked.

A similar phenomenon has also been observed with judgments of sarcasm [1]. Utterances dripping with sarcasm were filtered to mask lexical content while retaining prosodic information. Listeners rated the filtered sarcastic utterances to be more sarcastic than non-sarcastic utterances, as well as angrier and more inquisitive. What is not answered by these measures is what listeners’ judgments are when their attention is not focused on sarcasm, anger, or inquisitiveness. If listeners were simply asked what the response conveyed, would they infer different things? Do experimenters decrease the informativeness of their participants’ responses with their questions?

The methodology in the present study attempts to circumvent the issue of an experimenter’s unintentionally influencing and over-simplifying potentially sensitive judgments of behavior, specifically of personality gleaned from gesture. We tested what personality traits were conveyed by an animated agent by using both open-ended and closed, Likert scale questions. One way to characterize the two methods is that one assesses first impressions and the other assesses potentially primed impressions. Ideally, data obtained through open- and closed-ended questions should be complementary with open-ended question data enhancing the basic data provided by the scale inventory.

1.1 Seeing Personality

Visibility of a trait plays a role in how well a given trait is judged [6, 9]. For example, for extraversion, visibility is high and judgments are more accurate than for other personality traits [6, 9]. Extraverts also tend to consistently display broader gestures, made further away from their bodies [12, 18]. Their gestures are

also more frequent and animated than the gestures of introverts [11]. Findings such as these have been applied to agents designed to portray extraversion [15].

In contrast, emotional stability (i.e., neuroticism) and agreeableness are less visible and are judged less accurately [9]. Nevertheless, there is some evidence that different non-verbal behavioral cues can predict personality. The Five Factor Nonverbal Personality Questionnaire (FF-NPQ) asks the examinee how likely they are to engage in a particular behavior, for example, riding a bucking horse [17]. The FF-NPQ and the NEO-PI-R have been found to reliably predict the Big Five factors to a similar extent, indicating nonverbal behavioral cues can also be used as predictive of personality. That is, the validity of both verbal measures (i.e. adjective checklists) and nonverbal measures (e.g., FF-NPQ) suggests one can also determine personality directly from nonverbal behavior practices.

1.2 Expressing Emotional Stability

Previous researchers summarized the findings on emotional stability most applicable to an animated agent and were able to show reliable variation of the trait through variation in language and through the use or absence of scratches and other self-touches (*self-adaptors*) [14]. The current work examines whether a user’s impression an agent’s emotional stability can be determined solely based on its gestures and whether this impression changes depending on how the user is asked to evaluate the agent.



Fig. 1: The Shaky agent (proposed low emotional stability), shown on the left, was designed to display a stiff posture with rigidly held arms, raised shoulders and a narrow stance, whereas the Smooth agent (proposed high emotional stability) on the right was designed to appear more relaxed.

Based on the findings reported in Neff et al. [14], two new animation clips were generated, designed to portray high and low emotional stability. Both clips consisted of the same gestures with the same timing. Variation was only allowed in the quality of the movement. In brief, the high emotional stability variant was designed to appear more relaxed and comfortable, with gestures whose trajectories were smooth (Smooth Agent); in contrast, the low emotional stability

variant had jerky gestures (Shaky Agent). Motion capture data was used for both clips, with edits applied on top to vary qualitative aspects of the motion. The changes used are summarized in Table 1 and frames from the two clips illustrating the base pose difference are shown in Figure 1.

Table 1: Motion edits applied to create the Shaky Agent and Smooth Agent clips.

	Shaky Agent (Proposed Low Emotional Stability)	Smooth Agent (Proposed High Emotional Stability)
Stance	Narrower, asymmetric stance. Legs swiveled in.	Reduced knee bending.
Posture Timing	Time warped body turns to create more rapid jerks.	Used more smooth base motion data.
Collarbones	Collarbones brought up and more back.	Collarbones brought down and slightly back.
Arm Swivel	Elbow rotated 20° inward.	
Gesture Strokes	Gestures 20% smaller and brought more in front of body. Gestures cross in front of body. Jerks were added to the motion in the direction of the motion path.	Gestures more outward. No jerk was added to the motion, so the trajectory was smooth.
Gesture Retractions	Retract position is held out from the body, low and to the side of the character.	Hands are raised to allow arm bend and brought more in front of the body for a more relaxed appearance. Physical simulation is used on the retraction to add a more relaxed swing.
Gesture Phase Connections	Sharper, achieved by using lower weight tangents on motion curves.	More rounded.

2 Study

This study examined three questions. First, could gesture alone change an observers attribution of personality to a virtual agent? Second, was an observer’s unprimed attribution of an agents personality congruent with its proposed, programmed personality? Finally, were basic assessments of agent personality using human-normed Big Five inventories in agreement with the relatively unbiased impressions of personality? For the purposes of this study, we tested one personality dimension: emotional stability, as people often have a relatively strong folk notions on how they neurotic vs. non-neurotic individuals gesture.

2.1 Method

Participants. There were a total of 74 participants: 12 (16.2%) were recruited through UCSCs participant pool for class credit, 52 (70.3%) were Mechanical Turk workers for \$1, and 10 (13.5%) collected through convenience sampling of research assistants and their friends who were blind to the experiment. Thirty-five of the 74 participants completed a BFI inventory in addition to the single open-ended question.

Stimuli. Two 15-second clips were created using a single IVA with a covered face to avoid facial expression as a confound. He faced the participant and swayed at the identical times in both and gestured diagonally downwards from his shoulder with his arms. The agent that was programmed with proposed High Emotional Stability (non-neurotic) movement had smooth and sweeping (Smooth) movements, whereas the proposed Low Emotional Stability (neurotic) agent was characterized by jerky movements, reminiscent of someone who is shaking badly (Shaky).

Procedure. Prior to the start of the experiment, all participants were shown clips of recent video games to familiarize them with computer-generated animation that is not based on motion-capture, as our pilot data indicated that 10-15% of participants became preoccupied with how “unnatural” and “robotic” the virtual agent seemed. This habituation process increased the number of participants who were able to treat the agent as something that had a personality, though it did not reduce the number who provided a description about personality.

Using a between-subjects design to avoid carry-over effects, the participants watched either the Smooth or Shaky 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 given a modified version of the Big Five Inventory (BFI), a 44-item Big Five assessment that asks participants to rate the agent on perceived characteristics using 5-point Likert scales [8]. All items started with “If I had to guess, I would describe the character in the video as someone who” as opposed to the original “I am someone who” wording. They were not informed of the experimental aims until the end of the study.

2.2 Analysis

Open-ended question coding. Participants’ responses were coded using Goldberg’s Big Five clusters, which consist of 339 trait adjectives [7]. If a response was already listed, then the factor it was associated with would be counted as the personality factor that the participant found most salient. If a participant’s response was not found, two blind coders were asked to choose up to three of the closest Goldberg adjectives. All chosen adjectives had to apply to the same personality factor or they were excluded. For instance, the description

“he has leadership qualities” was excluded because one coder chose “dominant” (high extraversion) and another chose “cooperative” (high agreeableness). The descriptor “arrogant” was kept because the one coder chose “boastful” and the other, “pompous” (both low agreeableness). They also excluded descriptions that did not describe personality (e.g., “is gesturing”), as well as those that defied a readily apparent single adjective description (e.g., “like a small boy”). This was done because 1. only half of participants listed multiple adjectives that all converged upon a single personality factor (some responses applied to as many as four) and 2., the experimenters wanted to avoid over-ascribing meaning to descriptions that may not have been intended by the participants.

Big Five Inventory scoring. The BFI was scored according to John, Donahue, and Kentle [8] with personality scores calculated for all five dimensions: openness, conscientiousness, extraversion, agreeableness, and emotional stability (which they call “neuroticism”, so that high neuroticism is equivalent to low emotional stability).

2.3 Results

Open-ended question. Sixty-one (28 Smooth Agent, 33 Shaky Agent) participants produced 121 descriptors. Thirteen participants (17.6%) were dropped because they did not ascribe any personality to the agent. Thirty-seven descriptors were dropped from the analysis because they were not Big Five personality traits or were too ambiguous to fit into a single Goldberg (1990) factor cluster. Only 16 exact descriptors were found in the Goldberg (1990) 339-adjective inventory, so the remainder of descriptors had to be matched by the blind coders. Additionally, 30 participants listed adjectives that were all aligned with the same personality factor (half were single-word responses).

Table 2: Frequencies of Goldberg (1990) adjectives and adjective-equivalents produced by participants by factor.

	Openness		Conscientiousness		Extraversion		Agreeableness		Emotional Stability	
	High	Low	High	Low	High	Low	High	Low	High	Low
HES Agent	4	2	1	1	14	1	5	4	1	1
LES Agent	1	1	0	5	8	6	7	17	0	5

Fisher’s exact test revealed that participants’ spontaneous descriptions of a personality did differ depending on whether a participant watched a virtual agent designed to be either High Emotional Stability (Smooth) or Low Emotional Stability (Shaky) ($p = .01$). The modal personality trait by far for those observing the smooth agent was high extraversion while those observing the shaky agent

mentioned low agreeableness the most. 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 at all.

The open-ended question data suggested that the Smooth Agents gestures did not convey an emotionally stable personality but they did convey an extraverted personality. On the other hand, the Shaky Agent did seem to bear some gestural hallmarks of an emotionally unstable person, but most adjectives described him as having a disagreeable personality; that is, emotional stability was not the most salient trait conveyed.

Table 3: Structure matrix

	Function
Emotional Stability	0.90
Agreeableness	-0.74
Conscientiousness	-0.57
Extraversion	0.38
Openness	-0.29

Big Five Inventory. Examination of the BFI scale scores suggest that participants ascribed different personality traits to Smooth and Shaky Agents. T-tests revealed that there were significant differences between the agreeableness and the emotional stability ratings between the Smooth and Shaky Agents. The Shaky Agent ($M = 2.58$, $SD = 0.93$) was found to be significantly lower in agreeableness than the Smooth Agent ($M = 3.10$, $SD = 0.86$), $t(19) = -2.42$, $p = .02$, Cohen’s $d = 0.70$. He was also found to be less emotionally stable (or more “neurotic”) than the Smooth Agent, ($M = 3.42$, $SD = 0.95$ and $M = 2.81$, $SD = 0.80$ respectively), $t(72) = 2.92$, $p = .01$, Cohen’s $d = 0.69$. These results suggest that a difference in gesture alone can influence perceptions of personality as measured by the BFI.

A discriminant analysis was run in order to ascertain whether one agent’s gestures were better than the other in communicating personality. Echoing the t-tests, the structure matrix of correlations (see Table 3) suggested that the best predictor for distinguishing between Smooth and Shaky Agents was emotional stability, followed by agreeableness. The remaining BFI sub-scales are considered poor predictors. However, the agents were not equally effective in conveying personality. Classification results show a high rate of correct condition classification for participants who saw the agent that was meant to be low emotional stability, i.e., the Shaky Agent (86%). The success rate for the Smooth Agent participants was much lower (41.9%). Ideally, the patterns of BFI ratings would clearly indicate which video the participant watched, as the differences in gestures should

reliably convey two distinct personalities. However, only the Shaky Agent conveyed a specific personality in a more consistent fashion. Specifically, his gestures appeared to communicate a less agreeable and emotionally unstable personality.

2.4 Discussion

The results of this study show that the gestures and postures that a virtual agent is programmed to use can influence the perception of its personality, if people make the initial leap to ascribe a personality to it, which over 80% of the participants in this study did. The results also show that open-ended questions and Big Five inventories can sometimes yield conflicting information about what personality viewers ascribe to an IVA. The influence of gesture was not the same for both Smooth and Shaky Agents. Overall, the Smooth Agent’s gestures were somewhat unsuccessful in communicating a personality that was consistently identified by participants. The Shaky Agent, however, had his own distinct personality that came through from first impressions and in the BFI ratings.

The Shaky Agent was rated as being low in emotional stability using the BFI, which was the personality trait he was intended to evoke, though participants did not describe him as such in the open-ended question. On the other hand, he was also rated as being disagreeable on the BFI and spontaneously described as disagreeable in the open-ended question. This indicates that the participants’ first impressions of him were of a person who is hard to get along with socially, not because he is anxious or negativistic (traits of low emotional stability) but because he is “angry” and “someone of bad intent.”

Participants’ first impression of the Smooth Agent indicated that they thought he was highly extraverted, suggesting that the gestures he was programmed with primarily communicated extraversion rather than high emotional stability. This is not necessarily surprising as the outward trajectory of the arm movement has been shown to be related to extraversion [18]; there is no reason to believe one gesture cannot be related to more than one trait. Yet the BFI ratings showed that when participants were asked to speculate in a more directed fashion using a closed-ended question on his personality, he was rated no more extraverted than the Shaky Agent. He was, however, judged to be more agreeable and emotionally stable than the Shaky Agent.

3 General Discussion

In reality, gestures are often an accompaniment to language and people are almost never without some environmental or interactional context, all of which will weight interpretation of personality. Nevertheless, personality can be distinguished via gesture alone, even in a completely silent, faceless, decontextualized virtual agent. This suggests that gesture can be an effective way to subtly convey personality in an agent without having to provide any dialogue or backstory, though determining the gestures that go with specific personality traits may be more problematic.

Previous work done by Neff et al. [14] found that the manipulation of an agent’s non-communicative self-adaptor gestures (such as scratching) did influence perception of emotional stability. They also found that emotional stability and agreeableness were highly correlated linguistically but not gesturally; in fact, they found no effect of agreeableness for the non-verbal factors [14]. Our stimuli were completely non-verbal, slightly exaggerated versions of the Neff et al. stimuli. Differences in agreeableness were not intended or expected, although we did in fact find them.

It is possible that the nonverbal behaviors of agents are judged differently than those of humans, either because the variation in stylistic expression is viewed differently or because humans are reluctant to attribute personality traits based on an agent’s hand and body movements. It could also be that judgments of agreeableness and emotional stability are highly correlated, particularly if both are negative: for instance, if someone is disagreeable, one might be more prone to labeling him emotionally unstable and vice versa.

The BFI scores were useful in quickly rating an IVA’s personality, but it is only useful in comparison to another IVA. Both agents essentially scored neutral (between 2.6-3.6 range) ratings on all the BFI sub-scales: this is expected but not particularly informative. The analysis of the open-ended question was a more laborious process but provided data that gave a more vivid sense of what people thought of the agent’s personality, without priming them with the assumption that any Big Five traits were present or forcing those who did not think the agent had a personality to choose one for it. The BFI tells us that the Shaky Agent is less agreeable than the Smooth Agent, but agreeableness has many facets (warmth, generosity, empathy, temperament, etc.) and the BFI cannot narrow it down further. On the other hand, the open-ended responses tell us that he is disagreeable because he is most often described as “angry.”

Likewise, the most common responses for the Smooth Agent were “confused” and “open,” two words that do not converge on any Big Five categories. Only “open” (if indicative of high extraversion) could have been anticipated as a perceived trait *a priori* based on the research that created the stimuli (as outward gestures are also characteristic of emotional stability). Yet the same Smooth Agent who was rated as more emotionally stable than the Shaky Agent was also frequently described as being “confused.” This unanticipated result may have been overlooked if we had not given participants an open-ended question about their perception of the IVA’s personality.

The use of both qualitative and quantitative data in early stages of developing an IVA designed to evoke a certain personality can be useful in understanding a user’s perception. Future work will include a version of this study that uses gestures relating to extraversion, which is thought to manifest more physically than the other factors.

Acknowledgments. This research was supported by NSF grant IIS-1115742. We would like to thank our research assistants: Anna Wartan, Andrew Banakus, Jasmine Norman, Nicolas Russell, and Megan Kostecka.

References

1. Bryant, G., Fox Tree, J.: Recognizing verbal irony in spontaneous speech. *Metaphor and symbol* 17, 99–119 (2002)
2. Christenfeld, N.: Choices from identical options. *Psychological Science* 6, 50–55 (1995)
3. Christenfeld, N., Creager, B.: Anxiety, alcohol, aphasia, and ums. *Journal of Personality and Social Psychology* 70(451-460) (1996)
4. Costa, P., McCrae, R.: 'normal' personality inventories in clinical assessment: General requirements and the potential for using the neo personality inventory: Reply. (1992)
5. Fox Tree, J., Schrock, J.: Basic meanings of you know and i mean. *Journal of Pragmatics* 34, 727–747 (200)
6. Funder, D., Dobroth, K.: Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology* 52, 409–418 (1987)
7. Goldberg, L.: An alternative "description of personality": the big-five factor structure. *Journal of Personality and Social Psychology* 59, 1216–1229 (1990)
8. John, O., Donahue, E., Kentle, R.: *The big five inventory—versions 4a and 54*. Institute of Personality and Social Research, Berkeley: University of California, Berkeley (1991)
9. John, O., Robins, R.: Determinants of interjudge agreement on personality traits: the big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of personality* 61, 521–551 (1993)
10. John, O., Srivastava, S.: *The Big Five Trait taxonomy: History, measurement, and theoretical perspectives*, pp. 102–138. University of California, Berkeley (1999)
11. LaFrance, B., Heisel, A., Beatty, M.: Is there empirical evidence for a nonverbal profile of extraversion?: A meta-analysis and critique of the literature. *Communication Monographs* 71, 38–48 (2004)
12. Lippa, R.: The nonverbal display and judgment of extraversion, masculinity, femininity, and gender diagnosticity: Model analysis. *Journal of Research in Personality* 32, 80–107 (1998)
13. McCrae, R., Costa, P., Busch, C.: Evaluating comprehensiveness in personality systems: The california qset and the fivefactor model. *Journal of Personality* 54, 430–446 (1986)
14. Neff, M., Toothman, N., Bowmani, R., Fox Tree, J., Walker, M.: Don't scratch! self-adaptors reflect emotional stability. In: Springer (ed.) *Intelligent Virtual Agents*. pp. 398–411 (2011)
15. Neff, M., Wang, Y., Abbott, R., Walker, M.: Evaluating the effect of gesture and language on personality perception in conversational agents. In: Springer (ed.) *Intelligent Virtual Agents*. pp. 222–235 (2010)
16. Norman, W.: Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology* 66, 574–583 (1963)
17. Paunonen, S.: Big five factors of personality and replicated predictions of behavior. *Journal of Personality and Social Psychology* 84, 411–424 (2003)
18. Riggio, R., Friedman, H.: Impression formation: the role of expressive behavior. *Journal of Personality and Social Psychology* 50, 421–427 (1986)
19. Tupes, E., Christal, R.: Recurrent personality factors based on trait ratings. Tech. rep., Lackland US Air Force (1961)