

Whom Are You Willing to Confide in or Listen to: A Cheerful or Serious Chatbot?

ABSTRACT

We present an intelligent virtual interviewer that engages with a user in a text-based conversation and automatically infers the user's psychological traits, such as personality. Here we investigate how the personality of a virtual interviewer influences a user's behavior from two perspectives: user's willingness to confide in, and listen to, a virtual interviewer. We have built two virtual interviewers with two distinct personalities and deployed them in a real-world recruiting event. We present findings from completed interviews with 316 actual job applicants. Notably, users are more willing to confide in and listen to a virtual interviewer with a serious, assertive personality. Moreover, users' personality traits, inferred from their chat text, influence their perception of a virtual interviewer, and their willingness to confide in and listen to a virtual interviewer. Finally, we discuss the implications of our work on building hyper-personalized, intelligent agents based on inferred user traits.

Author Keywords

Virtual interviewer; chatbot; personality analytics; human-machine trust; computer personality; individual differences.

INTRODUCTION

From human capital management to healthcare, interviewing is often used to obtain information from target respondents and to assess certain characteristics of the respondents. For example, a hiring manager may interview a pool of job applicants over the phone to discover their career interests and assess their characteristics, such as personality and motivations, before determining their fitness for the position.

While traditional interviewing by people helps gather more and deeper information, it presents several limitations. First, it is difficult to scale since a human interviewer can afford to conduct only a limited number of interviews per day. Second, much research shows that a human interviewer's personal factors, such as personality, mood, or biases, can affect interview results (e.g., [19, 48]). Third, respondents may not feel comfortable disclosing sensitive information to human interviewers due to social desirability biases [27].

To overcome the limitations of a human interviewer while providing a personal touch, we are building a *virtual interviewer*, an intelligent agent that interviews a user through a text-based conversation. We call such a virtual interviewer a REP (Responsible, Empathetic Persona). As shown in Figure 1, not only does a REP (Kaya) conduct a text-based

virtual interview with a user (James), but it also *automatically* infers the user's traits, such as personality, based on the user's chat text given during the interview (see our submitted video demo). Our decision on supporting a text-only conversation is based on two reasons. First, compared to the use of richer modalities, such as speech or gestures, a text-only interaction can be supported much more robustly, which in turn suits our goal of deploying a REP for real world use. Moreover, text-only interaction keeps a user focused on the interview without being distracted [22, 38].

Compared to a human interviewer, a REP has two distinct advantages. First, a REP does not experience fatigue and can scale to interview hundreds of thousands of users simultaneously. Second, a REP can render a more objective, unbiased assessment, since it does not bring personal emotions or human prejudice into an interview.

Since studies show that interviewers' personalities can directly affect interview outcome [19, 48], one of our challenges is to equip a REP with the "personality" of an effective virtual interviewer. This challenge is non-trivial for three reasons. First, existing research suggests that computers should take on *one* strong, consistent personality [37]. However, it is unclear which personality a REP should assume to achieve multiple interview goals. For example, a warm and cheerful personality may aid a REP in building rapport with its respondents, who can then relax and act authentically [2, 27]. On the other hand, a more serious and assertive REP may elicit more respect and better cooperation from the respondents [19, 36]. Second, most existing systems stress the use of both verbal and non-verbal cues to express an agent's personality [25, 39, 36]. However, it is unclear how a REP can effectively exhibit its personality through text-only expressions. Third, unlike most systems, which are used in a laboratory setting, our REP is used in

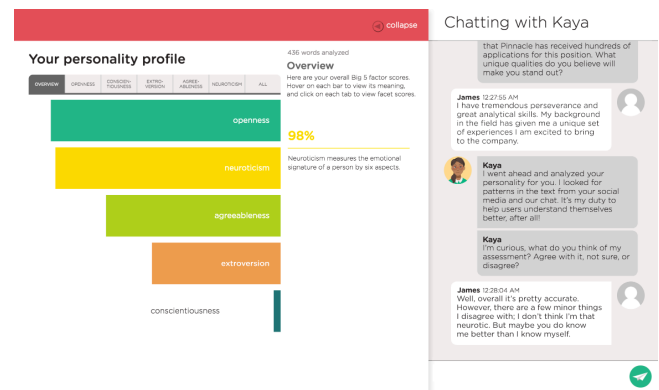


Figure 1. Our text-based chat interface during an interview.

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

the real world, including risk-bearing situations, such as job interviews. It is unknown how a REP’s personality would affect users in such a situation.

To address the challenge mentioned above, as the first step, we decided to give each REP a strong, consistent personality as suggested by existing work [37]. We thus built two REPs with two distinct personalities, each aimed at achieving a key interview goal. One is a female persona (Kaya) with a warm and cheerful personality, while the other is a male persona (Albert) with a reserved and assertive personality. Based on existing work, we hypothesize that:

- **H1:** Users are more willing to confide in Kaya since she is better at building rapport with them;
- **H2:** Users are more willing to listen to Albert and follow his advice since he is better at gaining respect from them;
- **H3:** Individual differences of users, including their personality, impact their behavior with Kaya or Albert.

We deployed both Kaya and Albert to help an organization interview and pre-screen their job applicants. In this field deployment, 157 applicants finished their interview with Kaya, and 159 completed theirs with Albert. Our results reveal several new insights. First, the personality of a REP affects users’ willingness to confide in and listen to the REP. Second, users’ individual differences, in particular, their personality traits inferred by a REP, influence their perception of and behavior with the REP. As a result, our work presents two unique contributions: (1) novel, practical approaches to building effective virtual interviewers who can automatically assess users’ characteristics; and (2) design implications for building hyper-personalized REPs based on our findings and automatically inferred user traits.

RELATED WORK

Our current work is influenced by studies on personification of agents and how the “personality” of machines influences user behavior. For example, studies show that equipping agents with human forms, such as faces and facial expressions, make the agents more likable and engaging [22, 38]. However, studies also find that it takes users’ effort to interpret the meanings of human-like expressions [22], which may even create a distraction [46]. Moreover, users are less relaxed or assured when interacting with more human-like agents [38]. As one of our main goals is to create an atmosphere where users can relax and act authentically during a virtual interview, we decide to support a text-only conversation to minimize potential user distraction from interpreting a REP’s human-like expressions.

To understand how the personality of a computer may influence user behavior, there is much work on matching agent personality with users’ personality and tasks. For example, Reeves and Nass [37] find that users favor a computer with a personality that matches their own. However, in human-robot interaction, studies show conflicting results. For example, Lee et al. [25] indicate that users enjoy interacting with a robot with a complementary personality,

while Tapus et al. [39] show that users prefer to interact with a robot with a matching personality in a rehabilitation therapy. Moreover, studies find that users expect the personality of an agent to match the task context, e.g., a serious agent for a serious task [36], which directly influences users’ willingness to comply with the robot’s instructions. While these studies provide insights into the effects of an agent’s personality on users in various laboratory situations, it is unknown what personality an effective virtual interviewer should take on and how its personality would impact its users, especially in real-world, risk-bearing situations.

Another line of related work is a wide variety of efforts on developing embodied “virtual humans” (e.g., [2, 11, 20, 27, 27]). Closest to our work are efforts on creating 3D virtual interviewers or counselors who build rapport with users to encourage the disclosure of sensitive information [2, 27]. Although less relevant, there are also virtual interviewers who use affective and social cues to build connections with users and teach them interview skills [11, 20].

While we leverage findings in these works, such as the use of introductory questions [27] and reciprocal self-disclosure [2] to build rapport, and the use of different virtual personas to influence users’ behavior [11, 20], our work differs in several aspects. First, although all these systems interpret user interaction behavior and maintain a user model, few focus on inferring a user’s psychological traits as ours does. Second, most existing efforts focus on instrumenting the social skills of an embodied agent to build connections with users. In contrast, our work aims at equipping a REP with a strong, consistent personality via text communication to achieve multiple interview goals, such as eliciting authentic user input and cooperation.

Our work on automatically inferring a user’s psychological characteristics, such as personality, from the user’s communication text, is inspired by multiple efforts on studying the relationships between users’ traits and their communication patterns. For example, research indicates that words or word categories in users’ blogs and essays are correlated with self-reported or third-party-rated personality traits [30, 40, 47]. More recent studies show that users’ vocabulary and behavior on social networks, such as Facebook, are also related to self-reported personality traits [21, 23]. Based on such findings, researchers have built various computational models to automatically infer one’s personality traits from text (e.g., [1, 14, 29]). Similar to these computational models, ours also automatically infers a user’s characteristics from text. However, unlike most of these models, which infer user traits based on the LIWC categories of word uses (e.g., [14, 29]) or high-level features (e.g., number of friends on Facebook [1]), our finer-grained computational model uses rich linguistic cues (e.g., words, phrases, emoticons, and punctuations) to infer user traits.

SYSTEM OVERVIEW AND KEY COMPONENTS

Figure 2 shows an overview of our virtual interviewer system. A REP conducts a virtual interview with a user (inter-

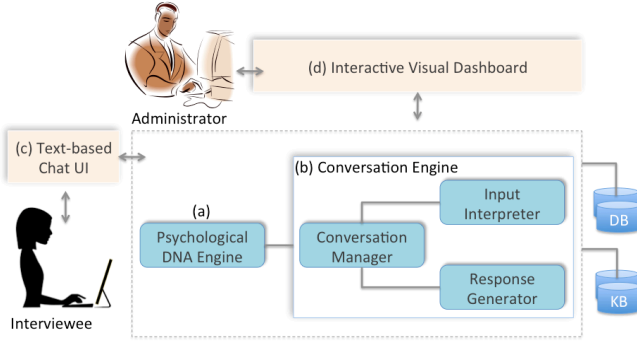


Figure 2. System overview.

viewee) through a text-based chat user interface (c), powered by two core engines: (a) psychological DNA engine and (b) conversation engine. The *psychological DNA engine* analyzes a user’s interaction behavior and automatically infers the user’s psychological characteristics, such as personality, strengths, and weaknesses. The *conversation engine* enables a REP to engage a user in a mixed-initiative conversation during a virtual interview. Similar to a human interviewer, a REP takes initiatives to guide the interview flow and solicit information from a user, while it allows the user to pose questions whenever appropriate. For example, a user may ask a clarification question or inquire about the position that s/he is applying for. Specifically, the *conversation manager* controls the conversation flow, the *input interpreter* processes user input, and the *response generator* outputs a REP’s verbal utterances.

To help business administrators (e.g., hiring managers) harness the interview results and make decisions, we provide an interactive visual dashboard that summarizes interview results (d). Figure 3 shows a sample of interview results. Our system also maintains several databases (DB) and knowledge bases (KB), which store various types of information, such as users’ answers to various interview questions and inferred user traits.

Our system uses a web-based, client-server architecture. The server-side components are written in Clojure, while the client ones are written in HTML and JavaScript.

Evidence-based Psychological DNA Engine

To automatically infer a user’s traits from text, most existing approaches use the uncovered correlations between word categories (i.e., LIWC categories) and self-reported personality test scores (e.g., [1, 14, 29]). While these approaches have demonstrated their effectiveness in various applications (e.g., [26, 28, 29]), they have three main limitations. First, while using word categories is simple, the inference is coarse. This is because different words belonging to the same word category are weighted the same but their relationships to a trait may be quite different [47]. Moreover, there is a limited number of word categories (68 in LIWC), and many word categories may not show significant correlations with a trait. Second, the correlations are

derived based on one’s word use and self-reported personality test scores, which may not be completely truthful [49]. Third, these approaches consider only single words but miss out other important linguistic cues, such as phrases, and punctuations, which also reflect one’s traits [21, 23].

To overcome these limitations, we have developed an evidence-based approach to inferring a user’s traits from the user’s text. While describing the approach in detail is out of the scope for this paper, we highlight the three main steps.

Step 1: Gathering Trait-Related Evidence

We first gather a person’s text that captures her real-world activities and reflects her associated characteristics. For example, a designer’s online portfolio captures her design activities in the real world, which reflects her level of *creativity*—a trait that we measure.

Step 2: Mining Evidence-Trait Relationships

We then mine the gathered data to extract various linguistic cues (evidence) and quantify their relationships with various traits. Linguistic cues may be words, phrases, punctuations, or emoticons, and serve as potential evidence to measure various traits.

To quantify the relationships between potential evidence and user traits, such as the Big 5 personality traits, we adopt a statistical inference approach, called Item Response Theory, the widely used approach in modern psychometrics [8]. In particular, we model the desired traits as latent factors and potential evidence as observed items. To estimate the relationships between any evidence j and a trait θ for an individual i , we use the following formula:

$$X_{ij} = u_j + \lambda_j \theta_i + \varepsilon_{ij}; j=1, N(1)$$

Here X is the observed data containing N pieces of evidence, u_j is the occurrence rate of evidence j , λ_j measures how discriminative evidence j to trait θ , and ε_{ij} is the Gaussian distributed error uncorrelated with θ .

To estimate the parameters in Formula 1, we use the conditional distribution of X for trait θ .

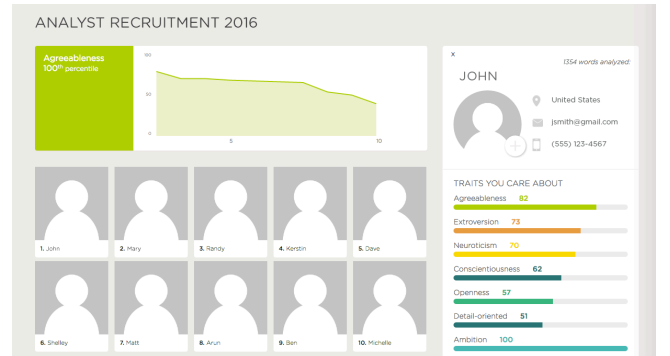


Figure 3. A sample interview result showing candidates ranked by REP-inferred Big 5 trait “Agreeableness”.

$$f(X_j|\bar{\mu}_i) = \frac{1}{\sigma_j\sqrt{2\pi}} \frac{1}{x_j(1-x_j)} e^{-\frac{1}{2}\left[\frac{\ln\left(\frac{x_j}{1-x_j}\right)-\bar{\mu}}{\sigma_j}\right]^2} \quad (2)$$

Here $\bar{\mu}$ is the expectation of μ after a logit transformation, and σ_j^2 is the variance of residuals. We estimate the parameters using an EM algorithm over the training data [4].

Step 3: Evidence-based Inference of Trait Score

Given a user’s text, this step analyzes the text and finds all matching linguistic evidence (e.g., matched words and phrases). The matching evidence is then used to compute a trait score by applying the trained model (Formula 1). Currently, we focus on inferring Big 5 personality traits, including the 30 associated facets [10]. Based on the inferred Big 5 traits, we also estimate relevant composite traits. For example, *Diligence* is composed of several Big 5 facets, such as *activity* and *achievement-striving*.

Topic-based Conversation Engine

To power virtual interviewers for practical, real world uses, our conversation engine must meet three design criteria. First, it must support diverse conversations and easy customizations of such conversations for a wide variety of interview situations, ranging from job interviews to patient inquiries to customer surveys. Second, the engine must be easily extensible to accommodate new capabilities, e.g., extending a virtual interviewer’s ability to coach a user. Third, the engine must allow easy integration of third-party conversational technologies, e.g., Google NLP functions.

By these three design criteria, and inspired by cognitive architectures [24], we have developed a novel, topic-based conversation engine to support mixed-initiative conversations. As described below, our approach models a conversation as a set of topics, which are flexibly threaded together to drive a conversation flow. As a result, a conversation can be easily customized to accommodate various interview situations through the specification of topics. Moreover, conversational capabilities can be easily extended, incorporating even third-party functions through topics.

Conversation Modeled as Topics

We model a conversation as a set of *topics*, each of which defines a specific subject of the conversation. Semantically, a topic is similar to a discourse segment [13]. Here are two example topics in a job interview conversation: one soliciting certain information from a candidate (Figure 4a) and the

other answering a candidate’s questions regarding the position (Figure 4b). Each topic further consists of one or more *semantic units*, each of which handles a particular aspect of the topic. For example, (*ask-user-intro*) has one unit (highlighted block), while (*answer-job-inquiry*) has two.

Each semantic unit is defined by a pair: a trigger and a response. A *trigger* encodes one or more conditions under which a semantic unit may be activated during a conversation, while a *response* defines one or more corresponding REP actions. In Figure 4(a), the condition to ask a user for a self-intro is at the beginning of an interview (*trigger* T1). If this condition is met, the REP asks the candidate to make a self-intro (*response* R1). As shown in topic (*answer-job-inquiry*), a trigger may be defined as a pattern that matches a user’s input, including a question like “*when a decision will be made*”. We describe below how a user’s input is matched with a pattern defined in a trigger. For extensibility, both trigger and response can contain patterns of expressions as well as functions. For example, response R1 (*ask-self-intro*) is a function that generates a question that asks a user to make a self-intro.

Optionally, a semantic unit may be associated with a *sub-topic* that lets a REP have a deeper conversation with a user on the topic. In Figure 4(a), (*drill-down*) is a sub-topic, which lets the REP probe further, e.g., asking the user to say more about herself. Sub-topics may be recursively defined to support virtually arbitrary levels of conversations.

Conversation Flow Driven by Topic Activation

Given a conversation that consists of a set of topics, our conversation engine decides when and which topics to activate, which in turn drives the conversation flow. To make these decisions, the conversation engine checks various properties of each topic. For this purpose, each topic is associated with a set of properties.

Initiator. First, there are three types of topics by their initiator. *Proactive* topics cover the subjects initiated by a REP in a conversation. In a virtual interview, proactive topics are often defined for a REP to pose interview questions and collect needed information from an interviewee. *Reactive* topics, on the other hand, cover the subjects initiated by a user. For example, (*ask-user-intro*) is a proactive topic, while (*answer-job-inquiry*) is a reactive one. We also support *mixed-initiative* topics, which may be initiated by a REP or by a user. For example, a chitchat topic may be initiated by a user (“*Shall we continue?*”), or by a REP (“*We still have time, let’s chat about your favorite food*”).

Structure. Sub-topics can only be activated if their associated parent topics are activated.

Importance. Just like in a human-human conversation, not all topics are equally important. Currently, we support three types of topics, ranked by importance. There are *agenda* topics that must be activated during a conversation at a certain point of time. In a virtual interview, agenda topics normally include all proactive topics that a REP wants to

| | |
|------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (deftopic ask-user-intro T1: [(chat-begin)] R1: [(ask-self-intro)] (drill-down)) | (deftopic answer-job-inquiry T2: [when make decision] R2: ["We will make a decision..."] T3: [how many apply] R3: [(answer-num-candidates)]) |
| (a) | (b) |

Figure 4. Topic representation. Labels T1-T3 and R1-R3 are not part of the script but added for the purpose of illustration.

bring up and discuss with a user. There are *chitchat* topics that may or may not be activated during a conversation. These topics are mostly used to help make conversational transitions or fill the gaps during a conversation. For example, assume that a REP has finished asking all the interview questions and there is still time left. One of chitchat topics may be activated to let the REP continue with a user. *Error-handling* topics are only activated if there are errors or unexpected situations occur during a conversation.

Temporal Order. Similar to a human-human conversation, certain topics (e.g., asking a user to introduce himself) are discussed before others (e.g., the user’s opinion on mobile strategy). A partial temporal order may be specified among defined topics to control the topic order. Topics may not be ordered at all to support an arbitrary flow among them.

To drive a conversation flow, the conversation engine examines all topics and their associated properties, and then decides which topic to activate. A topic becomes an activation candidate if one or more of its semantic units have matched triggers. If there are multiple candidates, the engine ranks them by their structure, importance, and temporal order. Specifically, it ranks an associated sub-topic the highest, then any agenda topics by their temporal order if any, followed by chitchat and error-handling topics. This ranking allows a conversation to go deeper *and* guarantees all agenda topics be discussed. By default, a semantic unit (e.g., asking a user to introduce himself) can be activated only once. After it is activated, the conversation about this aspect is considered done. However, certain semantic units are *reusable*. For example, all error-handling units are reusable, since they can be activated throughout a conversation to perform error handling whenever errors occur.

Pattern-based User Input Interpretation

Accurately interpreting a user’s input is one of the biggest challenges in building a conversational system. Although our REP infers a user’s traits based on observed linguistic cues (Formula 1) without the need to understand every user input precisely, it still must understand a user’s input adequately to drive the interview process. Given our goal of using a REP in the real world, we adopt a pattern-based approach to user input interpretation, since it has shown its effectiveness in many conversational systems [43, 45]. Each pattern is defined by a composition of tokens, strings, or regular expressions, which are recognized as states in a finite-state machine (FSM). For generalization purposes, where a pattern can match with many input variants, our engine automatically lemmatizes tokens and inserts wildcards. Patterns are compiled into optimized FSMs by minimizing the number of states [17]. The FSM compiler itself is optimized by using several novel pattern rewriting techniques, such as recursively factoring out common prefixes and suffices, retracting a large list of alternatives into a single special token, and eliminating duplicated patterns. As a result, hundreds of thousands of complex patterns can be compiled within seconds, and our FSM runtime can process

more than 100 million user input tokens per second on a mid-tier Macbook Pro Laptop.

Template-based System Response Generation

Currently, we use a template-based approach to generate a REP’s utterances during a conversation. To make a REP sound more natural, a response template may include multiple alternatives. For example, a response template that allows a REP to utter positive confirmations may include options, such as “*Great job*” and “*Well done*.” At run time, the conversation engine may select an option randomly or by certain criteria, e.g., linguistic cues reflecting a particular personality trait (e.g., *cheerfulness*).

DESIGN OF A VIRTUAL INTERVIEWER

To test our hypotheses, we developed two virtual interviewers, each with a distinct personality that is associated with effective human interviewers [19, 48].

Two Virtual Personas

Existing studies suggest two types of effective human interviewers: one with a warm, cheerful personality and effective at building rapport with interviewees, and the other with a serious, assertive personality and effective at requesting cooperation from interviewees [19, 48]. We thus designed and built two virtual personas.

Appearance

Since people often associate one’s appearance with a particular personality [22, 36, 46], we created two profile images, each representing a distinct persona. In particular, Kaya takes the form of a cheerful female, while Albert is a serious-looking male (Table 1). Our gender assignment is based on psychology studies, which show women in general are warmer and more sensitive in their personality, while men are more reserved and dominant [12, 44].

Conversational Behavior

In addition to appearance, an agent’s conversational behavior exhibits its personality [4]. Specifically, we unveil the personality of a virtual interviewer through four types of conversational behavior associated with an interview task.

Effective Inquiring. Clinical studies show that interviewers can use different techniques to effectively draw out truthful information from interviewees. For example, an *affective strategy* helps an interviewer establish rapport with an interviewee, which then allows the interviewee to open up and act authentically [15]. On the other hand, *negative politeness* may be used to preserve the emotional distance between an interviewer and an interviewee, minimizing imposition on the interviewee [5]. Since the use of particular strategies or techniques reveals an interviewer’s personality naturally, we match our virtual personas with different inquiring strategies or techniques. Specifically, aligning with her warm and cheerful personality, Kaya uses an affective strategy and positive politeness in a conversation, while Albert uses a cognitive strategy and negative politeness to reveal his reserved and serious personality.

Effective Influencing. Besides inquiring about information, an effective interviewer may want to influence interviewees (e.g., encouraging them to think outside the box) to draw out the best in them [33]. To achieve this goal, we introduce influence operators [4], which help our virtual interviewers effectively guide interviewees to perform their best. Similar to inquiring techniques, influence operators naturally exhibit an interviewer’s personality. We thus also assign different influence operators to our interviewers. Kaya uses empathetic and cooperative language to show her warm personality and influence, while Albert uses reassurance to exhibit his serious, assertive personality.

Small Talk. While small talk helps make a human-computer conversation more engaging [3], it may also block honest interaction [16]. To match their personalities, Kaya uses small talk to build rapport and show her friendly and cheerful side, while Albert refrains from chitchatting to establish his serious, authoritative persona.

Linguistic Style. Since one’s linguistic style unveils one’s personality [35], we incorporate matching linguistic cues (e.g., words, phrases, and punctuations) into our virtual interviewers’ utterances. These linguistic cues are selected based on previous studies [21, 23, 47] and our own findings that establish the relationships between various linguistic cues and personality traits (Formula 1). For example, Kaya frequently uses first-person, affective expressions (“*I love romantic movies as they make me cry.*”), emoticons, and exclamations to express her personality. In contrast, Albert uses declarative, projective statements in the third person (“*It makes sense.*”) or abbreviated replies (“*Understood.*”) to reflect his personality. Moreover, Kaya uses questions

and suggestions to show her friendliness, while Albert uses assertions and demands to display his seriousness. Table 1 summarizes the two REPs and their matched behavior.

Interview Questions

To test our hypotheses in a field deployment, we worked with a firm to screen their candidates for an open position. To satisfy the firm’s recruiting goals, we first collected a set of interview questions from them. Such questions were mostly about a candidate’s interests and skills pertinent to the job. We then added questions for three purposes: (1) making the virtual interview flow more naturally, (2) measuring candidates’ willingness to confide in and listen to a virtual interviewer, and (3) soliciting candidates’ perception and feedback of the virtual interviewer.

Overview of Interview Questions

All interview questions were organized into a five-part interview, which required about an hour to complete.

I. Introductory Section. The first section included five open-ended warm-up questions, helping a virtual interviewer build a rapport with a user. A REP started by asking the user to introduce herself, and then chatted with the user to collect additional information, such as her hobbies, favorite movie, and role model.

II. Impression Management Questionnaire. The second section included a 20-item impression management (IM) scale [34]. The IM scale, also known as a “lie” or social desirability scale, is designed to detect how users consciously favor themselves to impress others. We incorporated it to measure one’s willingness to confide in a REP.

III. Opinion Questions. This section included two open-ended opinion questions where users were asked to express their opinions first. Based on the users’ input, the REP persuaded them to read relevant articles and consider alternate viewpoints. Given the inherent controversial nature of the questions, these discussions were designed to measure users’ willingness to confide in and listen to a REP.

IV. About You. This section included 10 questions, asking users to expand about themselves, including their job preferences, skills and interests, and strengths and weaknesses. The REP also shared and discussed its analysis of a user, such as her personality traits, top strength, and top weakness. Based on such discussions, the REP also suggested certain actions, which were meant to measure the user’s willingness to listen to the REP.

V. Post Interview Survey. This section collected basic demographic information from users and asked the users to rate their impression of the REP on several aspects, from personality to performance.

Measuring Users’ Willingness to Confide in a REP

Drawing out truthful information from interviewees is paramount to any successful interview, since such information helps both interviewers and interviewees better assess real life situations (e.g., the fitness between a job and a candi-



| | Kaya | Albert |
|------------------------------|--------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| Profile |  |  |
| Personality | Gregarious, Cheerful, Warm, Agreeable, Humorous; Like a friend | Reserved, Calm, Assertive, Rational, Careful; Like a counselor |
| Effective Inquiring | Affective strategy; Positive politeness | Cognitive strategy; Negative politeness |
| Effective Influencing | Empathy, Comfort, Frankness, Cooperation, Agreement | Reassurance, Commitment, Forgiveness |
| Small Talk | Personable | Minimal chitchatting |
| Linguistic Style | Questions, suggestions, affective expressions. Example expressions: “!” 😊 “awesome”. | Assertions, projective statements, terse expressions. Example expressions: “evidence”, “determine”. |

Table 1. Two virtual personas and their behavior.

date) [33]. Due to social desirability biases especially in risk-bearing situations like job interviews, however, interviewees may not be completely truthful [34, 49]. While studies show that users are more willing to confide in machines [27], we want to further understand how the personality of a REP affects one’s willingness to confide in the REP, especially in a risk-bearing situation.

Based on previous studies [2, 15, 27], we hypothesize that users are more willing to confide in a REP with a friendly, cheerful personality (Hypothesis I). To test our hypothesis, we assess one’s willingness to confide in a REP by their willingness to share sensitive information with the REP during an interview. In particular, we measure such willingness through two sets of interview questions: the Impression Management (IM) scale, and three sharing actions of revealing sensitive information about oneself or one’s opinion on controversial topics with the recruiting firm.

The IM scale [34], described earlier, is designed to measure how much one consciously inflates herself to impress others. A user rates herself using a 7-point Likert scale on each of 20 items, such as “*I sometimes tell lies if I have to*”. The higher the score, the more inflation there is, and thus the less willing one is to confide in another with honest ratings.

In addition to the IM scale, we designed three user actions to measure one’s willingness to confide in a REP. Each action is related to sharing with the hiring manager certain sensitive information or an opinion of a controversial subject. Table 2 summarizes these three actions. The first was to share a user’s biggest weakness. The REP asked a user about her biggest weakness before showing its inferred top weakness for the user. The REP then asked the user to rate the inferred weakness by a *rating* in Table 2. It then asked the user to choose one of three actions (*sAction* in Table 2). The user’s rating and choice of action were used to compute the user’s willingness to confide in a REP (*A* in Table 2). By this measure, a user is most willing to confide in the REP if she chooses to share her biggest weakness suggested by the REP even if she *disagrees* with its analysis.

Two additional sharing actions concerned controversial subjects (e.g., “*What important truth do very few people agree with you on?*”). For each subject, the REP first asked a user’s opinion and then probed further by eliciting rationale or introducing counter viewpoints. The REP asked the user how confident he was (*CF* in Table 2) before asking him whether he still wanted to share his opinion with the hiring manager (*sAction* in Table 2). Based on the user’s reported confidence and choice of action, the willingness to confide in is then computed (Table 2). By our measure, a user is most willing to confide in a REP if he decides to share his opinions with the hiring manager despite having no confidence. Our rationale behind this measure is that users always want to show their best qualities in an interview. When they have an option not to share something they are uncertain of, they could hold back without risking

| | | | |
|---------------------------------------|------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|--|
| | | $A = rating \times sAction$ | |
| Share Weakness | $rating = \begin{cases} 3, disagree \\ 2, not\ sure \\ 1, agree \end{cases}$ | $sAction = \begin{cases} 0, don't\ share \\ 1, share\ mine \\ 2, share\ REP\ sugg. \end{cases}$ | |
| | | | |
| | | $B_j = CF_j \times sAction_j$ | |
| Share Opinions ($j=1, 2$) | $CF_j = \begin{cases} 1, High \\ 2, Med \\ 3, Low \end{cases}$ | $sAction_j = \begin{cases} 0, don't\ share \\ 1, share \end{cases}$ | |
| | | | |
| | | $Willingness\ to\ confide: WC = A + \sum_{j=1}^2 B_j$ | |

Table 2. Measuring willingness to confide in a REP.

to expose their potential weaknesses (unless they are willing to confide in a REP despite the risks).

Measuring Users’ Willingness to Listen to a REP

In addition to eliciting authentic responses from interviewees, an interviewer should also effectively guide the interviewees and help them reveal their best qualities (e.g., demonstrating thoughtfulness). However, users may not always be willing to listen to machines [36, 37]. To test our hypothesis II, where users may be more willing to listen to a REP with a serious personality [19, 36], we designed a set of user actions to measure one’s willingness to listen to a REP. Table 3 summarizes the actions and their measures.

In the first two actions, a REP asked a user to read specific online articles related to the current interview topic by clicking on the articles’ URLs. In these two cases, *willingness to listen* is measured by a user’s click-throughs. Another five actions were related to persuading a user to share a piece of REP-suggested information, often different from his originally expressed opinion, with the hiring manager. In these five actions, *willingness to listen* is measured by how strongly a user disagreed with a REP’s suggestion and the user’s final choice of action.

Similar to the classic Desert Survival Problem, which is often used to test the power of persuasion [37], all of our five actions on testing one’s willingness to listen followed a four-step process. The REP first asked a user’s input on a specific subject (e.g., her top strength) before offering a suggestion (e.g., her top strength inferred by the REP), which was often different from the user’s input. The REP then asked the user to rate its suggestion (*rating* in Table 3) before persuading her to share its suggestion, as opposed to the self-reported one, with the hiring manager.

Measuring Users’ Perceptions of a REP

Based on previous studies on anthropomorphic interfaces [25, 36, 38, 39] and our hypotheses, we designed three sets of Likert-scale questions to measure users’ perceptions of a REP. The first set asked users’ perceived trustworthiness of a REP, since such perceptions may be directly related to their willingness to confide in and listen to the REP. To validate the effectiveness of our personality manipulation (i.e., whether users perceived a REP’s given personality), the second set asked users’ perceptions of a REP’s personality from multiple aspects, such as perceived characteris-

tics (e.g., how *cheerful* or *reserved* a REP is), role similarity (i.e., how a REP is similar to a friend or a counselor), and personality similarity (i.e., how similar a user’s personality is to a REP’s). The third set solicited user’s feedback on a REP’s usability (e.g., how enjoyable a REP is) [37].

FIELD DEPLOYMENT

We deployed REP Kaya and REP Albert to aid an investment firm in hiring an analyst. The firm received about 800 applications for the position. They first notified all their applicants that they should expect an email invitation from us about a virtual interview opportunity and encouraged them to use this opportunity to showcase their unique characteristics and talents. Following their email, we invited 368 candidates via email to interview with Kaya, and 398 candidates to interview with Albert. Each email contained an individualized, unique URL to access our website and start the interview. The URL automatically expired once it was used, preventing multiple interview attempts.

The process lasted about three weeks, during which 157 applicants completed their virtual interviews with Kaya, 159 completed theirs with Albert, and 59 people started but did not finish their interviews. Among the 157 users with Kaya, there were 130 (83%) male and 27 female (17%); among the 159 users with Albert, there were 125 (79%) male and 33 (21%) female. Users’ ages ranged from over 18 to over 50, and 90% of them were between 18 to 35 years old. On average a user spent over an hour with a REP, contributing about 1600 words.

RESULTS

Since a unique capability of our REP is to assess a user’s traits automatically, we have evaluated our trait inference engine in several ways, although completely validating such a model is non-trivial [14]. First, we used our model to analyze over 14 million Twitter and Facebook users. We randomly selected 100,000 of them to evaluate the reliability of our model by computing the Cronbach’s α , a metric estimating the reliability of a psychometric test [7]. Our results showed that our model requires about 1000 words to achieve a good reliability ($\alpha \geq 0.8$) for many traits (e.g., 19 of 35 Big 5 personality traits). Second, about 400 users responded to our survey on rating the inferred traits of themselves and their friends. Of these, 91% of them thought their own inferred traits were accurate, while 75% thought the inferred traits of their friend were accurate. Third, our

current field deployment demonstrated the effectiveness of our model. Using our dashboard, the hiring manager examined top-50 candidates ranked by a set of inferred traits that he cared for (e.g., cautiousness). In conjunction with other criteria (e.g., skills listed in a resume), he chose 12 out of the 50 to follow up with a phone or onsite interview. Through the follow-up, he found all 12 a good fit for the position. He also chose 4 outside of the top 50 to follow up and found two of them were a good fit but two were not.

Overview of Results

We first present an overview of the results and then our analysis around our three hypotheses.

After each interview, we asked users to rate their perception of the REP’s personality traits (on a scale 1-5). The top three rated personality traits for Kaya were: *calm*=3.89, *cheerful*=3.875, and *warm*=3.82, and three for Albert were: *calm*=4.15, *rational*=3.96 and *assertive*=3.58. Except the highly rated *calm* for Kaya, users’ perceptions matched with the intended personality of Kaya and Albert (Table 1). One explanation for Kaya’s calm rating is that her role as a job interviewer might have given users such an impression.

An independent samples t-test also showed that users rated Kaya significantly more like a friend (mean=2.74, sd=1.20) compared to Albert (mean=2.39, sd=1.21), $t(307)=2.54$, $p < .01$. A trend also showed that users rated Kaya more likeable (mean=3.54, sd=1.12) than Albert (mean=3.32, sd=1.03), $t(311)=1.80$, $p < .07$. Each REP was also rated on three trust measures (integrity, competence, and benevolence) [31], which were combined additively to create an index measure called *Trust*. An independent samples t-test showed a strong trend that users trusted Albert (mean=3.84, sd=0.82) more than Kaya (mean=3.65, std=0.99) $t(314)=1.88$, $p < .06$. There were no significant differences in other user ratings, such as finding a REP helpful.

H1: User’s Willingness to Confide in a REP

Our first hypothesis stated that users would be more willing to confide in Kaya as she should be better at building rapport. As mentioned earlier, we measured one’s willingness to confide in a REP from two aspects: the IM Scale [34] and three actions measured by an index variable *Willingness to Confide* (Table 2).

By the IM Scale, users were more honest with Albert (mean=12.68, sd=3.48) than with Kaya (mean=12.93, sd=3.44). Using the IM scale as a dependent variable, we ran a General Linear Model (GLM) with three independent variables: *Trust* (user perceived trustworthiness of a REP), *SimPersonality* (how similar users felt the REP’s personality was to their own), and *Agent* (distinguishing Kaya from Albert), controlling for user’s *Gender*. The result showed *Trust* was the only significant factor: $F(1, 302)=5.65$, $p < .02$, $Coeff=.59$. This implies that the more users trust a REP, the more likely they want to consciously impress the REP. This finding seems consistent with the theory that people apply the same social rules to computers as they

| | |
|---------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Click-through ($j=1, 2$) | $C_j = \begin{cases} 0, & \text{URL not clicked} \\ 1, & \text{URL clicked} \end{cases}$ |
| $S_j = \text{rating} \times \text{act}$ | |
| Share Action ($j=1, 5$) | $\text{rating} = \begin{cases} 3, & \text{disagree} \\ 2, & \text{not sure} \\ 1, & \text{agree} \end{cases} \quad \text{act} = \begin{cases} 1, & \text{share REP's} \\ 0, & \text{share mine} \end{cases}$ |
| $\text{Willingness to listen: } WL = \sum_{j=1}^2 C_j + \sum_{j=1}^5 S_j$ | |

Table 3. Measuring willingness to listen to a REP.

apply to humans [37]. Specifically, people want to impress someone whom they trust (e.g., teachers and parents).

With *Willingness to Confide* as a dependent variable, we next ran a GLM with the same three independent variables: *Trust*, *SimPersonality*, and *Agent*, also controlling for *Gender*. It showed that *Agent* was significant: $F(1, 302)=5.38$, $p<.02$, $Coeff=.46$. *Trust* showed a trend: $F(1, 302)=3.22$, $p<.07$, $Coeff=-.22$. This shows that users are more willing to confide in Albert and they are more willing to do so if they trust the REP more. *Gender* was not significant, indicating that the gender of the user did not influence users' willingness to confide in a REP.

Our results above show that users are more honest (per the IM) and more willing to confide in Albert, which is *contrary* to our hypothesis that users would do so more with Kaya.

H2: User's Willingness to Listen to a REP

Our second hypothesis assumed that users would be more willing to listen to Albert, who should be better at establishing authority. Our dependent variable was an additive index of users' actions, *Willingness to Listen* (Table 3).

With *Willingness to Listen* as a dependent variable, we ran a GLM over the same independent variables: *Trust*, *SimPersonality*, and *Agent*, controlling for *Gender*. It showed that *Agent* was significant: $F(1, 302)=4.15$, $p<.04$, $Coeff=.39$, and that *SimPersonality* showed a weak trend: $F(1, 302)=2.98$, $p<.09$, $Coeff=.17$. This result showed that users are more willing to listen to Albert, a REP with a reserved, assertive personality. Moreover, users might be more willing to listen to a REP who possesses a similar personality, consistent with certain previous findings [37, 39].

H3: Effects of Individual Differences on User Behavior

In our third hypothesis, we expected that users' individual differences, i.e., their traits inferred from their text during a virtual interview, would affect their perceptions of and behavior with a REP. A factor analysis was done first on the 35 inferred Big 5 personality traits [10]. We used a Varimax rotation with a Kaiser normalization. A scree plot¹ revealed that seven factors should be used, accounting for 50.5% of the variance (Table 4). To test our hypothesis, we used these factors as independent variables in regression models below.

Personality Traits and User Perception of a REP

We analyzed how the inferred users' personality traits may affect their perception of a REP. We focused on two user perception variables that were most relevant to a virtual interview context: *Trust* and *Liking* a REP. Per REP, we first conducted a stepwise regression using *Trust* as the dependent variable and the seven factors of personality traits as independent variables. We found that for Albert, *Factor 4* significantly impacted *Trust* $F(1, 153)=6.83$, $p<.01$, $Coeff=.17$, whereas for Kaya, *Factors 4* and *5* significantly

affected *Trust* $F(1, 154)=7.58$, $p<.001$, $Coeff$ (*Factor 4*)=.15, $Coeff$ (*Factor 5*)=.25.

We next investigated how the personality factors corresponded with users' liking a REP. A stepwise regression with the seven trait factors showed that for Albert, *Factors 3* and *4* significantly impact users' liking Albert $F(1, 151)=4.40$, $p<.01$, $Coeff$ (*Factor 3*)= -.16, $Coeff$ (*Factor 4*)=.18, while no significant personality trait factors were found related to liking Kaya.

Our results show that conscientious (*achievement-striving* and self-discipline are facets of *conscientiousness*) users expressed that they like Albert and trust a REP in general.

Personality Traits and User Behavior with a REP

In addition to analyzing the effects of the inferred traits on users' perceptions of a REP, we examined how such traits affect users' willingness to confide in and listen to a REP.

Per REP, we conducted a stepwise regression with *Willingness to Confide* as the dependent variable and the seven factors as independent variables. *Factor 1* was significant: $F(1, 155)=4.19$, $p<.04$; $Coeff = -.26$, showing that more emotional, worried, and vulnerable users are *less* willing to confide in Kaya. However, no significant trait factors were found for Albert.

We also conducted a stepwise regression with the *IM* scale as the dependent variable and the seven factors as independent variables. *Factors 2* and *3* were significant for Albert: $F(2, 152)=4.87$, $p<.009$; $Coeff$ (*Factor 2*) =-0.59, $Coeff$ (*Factor 3*)=-0.63; while no significant traits were found for Kaya. As a higher value of IM indicates a greater tendency to inflate one's own image, the results show an inverse relationship: extroverted and emotional users are less likely to inflate themselves to impress a serious REP.

Finally, we conducted a stepwise regression with *Willingness to Listen* as the dependent variable, and the seven factors as independent variables. *Factor 7* was significant for Kaya: $F(1, 155)=4.97$, $p<.03$; $Coeff=.28$, implying that more agreeable and trusting users are more willing to listen to a warm and cheerful persona like Kaya. However, no trait factors were found significant for Albert.

Summary of Findings

We summarize three key findings and their implications on designing effective virtual interviewers.

First, our field deployment demonstrated several practical values of a virtual interviewer. It makes a recruiting process

| |
|------------------------------------------------------------------------------|
| Factor 1: Emotionality, Anxiety, Vulnerability |
| Factor 2: Extraversion ← Cheerfulness |
| Factor 3: Neuroticism ← Depression |
| Factor 4: Self-Discipline → Conscientiousness ← Achievement-Striving, |
| Factor 5: Adventurousness → Openness |
| Factor 6: Intellect, Liberalism |
| Factor 7: Agreeableness ← Trust |

Table 4. Personality traits loaded onto separate factors. The arrow indicates one trait is a facet of another.

¹ A scree plot is used to determine the number of factors to select based on a visual analysis of when the curve flattens.

more efficient, objective, and inclusive. Out of 800 applicants, the hiring manager was able to quickly identify 12 quality candidates. He also commented that two of 12 would have gone unnoticed if judging their resumes alone. Moreover, users act authentically around a virtual interviewer to reveal their true character. For example, one told Albert that his question was dumb, while another told Kaya to “drink warm water” after she said “I just had a brain freeze” when recovering from an error. Users also “felt more casual speaking with Kaya than an actual person, almost like writing in a journal”, while others found “[REP’s analysis] very insightful... It was not something I had thought about before and appreciate the information”.

Second, create a virtual interviewer with Albert’s personality and its matching behavior (Table 1). This makes users willing to confide in and listen to the interviewer.

Third, it is desirable to customize a virtual interviewer’s behavior based on a user’s inferred traits to make an interview more valuable. For example, our results show that achievement-striving users tend to trust a REP more and such trust also makes them eager to impress the REP. In such a case, a REP can remind such users to be honest because they can achieve more with their true selves.

DISCUSSION

Here we discuss the limitations in our current work, and its design implications on building hyper-personalized intelligent agents based on automatically inferred users’ traits.

Limitations

Our current work has several limitations due to its experimental context and its current implementation.

Experimental Constraints

While we have tested our hypotheses in real-world job interviews, this context posed a couple of limitations on our understanding of the results. First, our users were real job applicants who had much at stake; thus the risk-bearing situation itself might have influenced their interview behavior. It is unknown whether users would behave similarly in a less risky context, such as routine patient interviews or customer surveys. Second, while previous work shows that gender impacts user interaction with agents [38], we found that user’s gender did not affect their interactions. One explanation is that the 60 female users were not a representative sample of the female population, which could have affected the results. As we are deploying REPs to support diverse interview situations, it would be interesting to see whether our findings would hold across interview contexts.

Customizing an Interview

Currently, each interview question is represented by a conversation topic, and follow-up questions are scripted as sub-topics. While it is straightforward to customize an interview by adding or removing interview questions as topics, it is impractical for a business user, such as a hiring manager, to do so manually. The task becomes even harder if an interview question is an open-ended question, which would then

require a business user to consider possible user responses and how to handle them as sub-topics. Ideally, a business user needs to enter just her interview questions, which are then automatically translated into conversation topics. Toward this goal, we are examining patterns in interview questions before auto-translating them into topics.

Inferring User Traits Beyond Text

Currently, a REP infers a user’s traits from the user’s text given during the interview. However, we observed interesting user behavior beyond what their text captures during the whole process. For example, certain users asked for help (via email) when encountering technical difficulties (e.g., using an incompatible web browser) while others simply gave up (e.g., quite a few abandoned partially completed interviews). Moreover, during an interview, a user may respond to a question promptly but hesitate to reply to another question. In addition to a user’s text, all such behavior could have been leveraged to build a better trait inference model. The challenge is how to capture such behavior, especially outside of a virtual interview, and how to relate user behavior with specific traits.

Design Implications Beyond Virtual Interviewers

Our findings indicate that inferred user traits significantly influence users’ behavior with a REP. This implies that a REP can dynamically adapt its behavior to a user based on the user’s traits inferred during a conversation. For example, when an *impulsive* user is unwilling to listen to a REP’s guidance during a coaching session, the REP would try to calm him down first before continuing. On the other hand, a REP would encourage a *humble* person to open up and show her best talents. While much research recognizes the importance of incorporating user traits, such as personality, into an intelligent system [18, 41], few practical systems have been developed to obtain and use such traits in real time. Thus, automatically inferring user traits during a human-computer conversation enables the development of a new generation of intelligent agents, who can truly understand their users on the fly and adapt to them in real time.

CONCLUSIONS

We have presented an intelligent virtual interviewer, called a REP, who interviews a user through a text-based conversation and automatically infers the user’s traits, such as personality and strengths. In this paper, we show how two REPs with distinctly different personalities influence users’ behavior—their willingness to confide in and listen to a REP during an interview. Our results from a field deployment of the two REPs with 316 completed interviews revealed several new findings. First, the users are more willing to confide in and listen to a serious, assertive REP. Second, certain personality traits of users, influence their perception of and behavior with a REP. Our findings imply that we can create hyper-personalized, effective REPs based on users’ traits, which can be automatically inferred during a conversation.

REFERENCES

1. Adali, S., & Golbeck, J. Predicting personality with social behavior. *ASONAM'2012*, 302–309.
2. Bickmore, T., Gruber, A., & Picard, R. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, 2005, 59(1): 21–30.
3. Bickmore, T. & Cassell, J. Small talk and conversational storytelling in embodied interface agents. *AAAI Fall Symposium on Narrative Intelligence*, 1999.
4. Bouchet, F. & Sansonnet, J. Intelligent Agents with Personality: from Adjectives to Behavioral Schemes. In E. M. Alkhalifa & K. Gaid (Eds.), *Cognitively Informed Interfaces: System Design and Development*, 2012, 177–200. IGI Global.
5. Brown, P. & Levinson, S. *Politeness: Some universals in language usage*. Cambridge University Press, 1987.
6. Cialdini R. *Influence: The Psychology of Persuasion*. 2006. Harper Business.
7. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, 16 (3): 297–334.
8. de Ayala, R. J. *The Theory and Practice of Item Response Theory*, Guilford Press. 2009.
9. Dempster, A., Laird, N., & Rubin, D. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Society*, 39(1): 1–38, 1977.
10. Digman, J. M. Personality structure: Emergence of the five-factor model. *Annual Review of Psychology* 1990, 41(1): 417–440.
11. Gebhard, P., Baur, T., Damian, I., Mehlmann, G., Wagner, J. & André, E. Exploring interaction strategies for virtual characters to induce stress in simulated job interviews. *AAMAS 2014*: 661–668.
12. Gludice, M., Booth, T., & Irwing, P. The distance between mars and venus: Measuring global sex differences in personality. *Pub. Library of Sci.* 2012, 7(1).
13. Grosz, B. & Sidner, C. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3): 175–204. 1986.
14. Gou, L., Zhou, M.X., & Yang, H. KnowMe and ShareMe: Understanding automatically discovered personality traits from social media and user sharing preferences. *CHI '14*, 955–964.
15. Havens L. *Making Contact: Uses of Language in Psychotherapy*, Harvard University Press, 1986.
16. Helgoe, L. *Introvert Power: Why Your Inner Life Is Your Hidden Strength*. Sourcebooks, 2013, 2nd Ed.
17. Hopcroft, J., An nlogn algorithm for minimizing states in a finite automaton, *Theory of machines and computations*, Academic Press, 189–196, 1971.
18. Hu, R., Pu, P.: Enhancing collaborative filtering systems with personality information. *RecSys '2011*
19. Jackle, A., Lynn, P., Sinibaldi, J., & Tipping, S. The effect of interviewer experience, attitudes, personality, and skills on respondent cooperation with face-to-face surveys. *Survey Research Methods*, 2013, 7(1): 1–15.
20. Jones, H., Sabouret, N., Damian, I., Baur, T., André, E., Porayska-Pomsta, K., & Rizzo, P. Interpreting social cues to generate credible affective reactions of virtual job interviewers. CoRR abs/1402.5039 (2014).
21. Kern, M., Eichstaedt, J., Schwartz, A., Dziurzynski, L., Ungar, L., Stillwell D., Kosinski, M., Ramones, S., & Seligman M. The online social self: An open vocabulary approach to personality. *J. of Assessment*, 2013, 21(2), 158–169.
22. Koda, T. & Maes, P. Agents with faces: The effects of personification of agents. *Proc. HCI'96*, 98–103.
23. Kosinski, M., Stillwell, D., & Graepel, T. Private traits and attributes are predictable from digital records of human behavior. *Proc. National Academy of Sciences of the U.S.A.* 2013, 110(15), 5802–5805.
24. Laird, J.E. *The Soar Cognitive Architecture*, MIT Press, 2012.
25. Lee, K., Peng, W., Jin, S., & Yan, C. Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human-robot interaction. *J. Comm.*, 2006, 56, 754–772.
26. Lee, K., Mahmud, J., Chen, J., and Zhou, M.X. & Nichols, J. Who will retweet this? Automatically Identifying and Engaging Strangers on Twitter to Spread Information. *IUI 2014*: 247–256.
27. Lucas, G., Gratch, J., King, A., & Morency, L. It's only a computer: virtual humans increase willingness to disclose. *Comp. in Human Behavior*, 2014, 37: 94–100.
28. Luo, L., Wang, F., Zhou, M.X., Pan, Y. & Chen, H. Who have got answers? Growing the pool of answerers in a smart enterprise social QA system. *IUI'2014*, 7–16.
29. Mahmud, J., Zhou, M.X., Megiddo, N., Nichols, J. & Drews, C. Recommending targeted strangers from whom to solicit information on social media. *IUI 2013*: 37–48.
30. Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. Using linguistic cues for the automatic recognition of personality in conversation and text. *JAIR*, 2007, 30: 457–500.
31. Mayer, R. C., Davis, J. H., & Schoorman, F. D. 1995. An integrative model of organizational trust. *Academy of Management Review*, 709–734.
32. Mehlmann, G., Janowski, K., & André, E. Modeling Grounding for Interactive Social Companions. *KI* 2016, 30(1): 45–52.

33. Okun, B. *Effective Helping: Interviewing and Counseling Techniques*. 7th Edition, Cengage Learning, 2007.
34. Paulhus, D. L. Measurement and control of response bias. In J. P. Robinson & P. R. Shaver (Eds.), *Measures of personality and social psychological attitudes* 1991, vol. 1, 17–59. Academic Press.
35. Pennebaker, J. & King, L. Linguistic styles: Language use as an individual difference. *J. Personality Social Psychology*, 1999, 77(6): 1296-1312.
36. Powers, A., Kiesler, S., & Goetz, J. Matching robot appearance and behavior to tasks to improve human-robot cooperation. The 12th IEEE Intl. Workshop on Robot and Human Interactive Communication, 2003, vol., IXX, 55-60.
37. Reeves, B. & Nass, C. *The Media Equation*. CSLI Publications, 2002.
38. Sproull, L., Subramani, M., Kiesler, S., Walker, J., & Waters, K. When the interface is a face. *J. Human-Computer Interaction*, 1996, vol. 11, 97-124.
39. Tapus, A., Tapus, C., & Mataric, M. User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics*, 2008, 1:169–183.
40. Tausczik, Y. & Pennebaker, J. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *J. of Language and Social Psychology*, 2010, 29(1): 24–54.
41. Tintarev, N, Dennis, M., & Masthoff, J. Adapting Recommendation Diversity to Openness to Experience: A Study of Human Behavior. *UMAP 2013*: 190-202.
42. Uziel, L. Impression management (“lie”) scales are associated with interpersonally oriented self-control, not other deception. *Journal of Personality*, 2013, 82(3): 200-212.
43. Wallace, R., Tomabechei, H., & Aimless, D. Chatterbots Go Native: Considerations for an eco-system fostering the development of artificial life forms in a human world. <http://www.pandorabots.com/pandora/pics/chatterbotsgonative.doc>. 2007.
44. Weisberg, Y., DeYoung, C., & Hirsh, J. Gender differences in personality across the ten aspects of Big Five. *J. Front Psychology*, 2011, 2:178.
45. Weizenbaum, J. ELIZA – a computer program for the study of natural language communication between man and machine. *Comm. of the ACM*, 1966, 9:36–45.
46. Wexelblat, A. Don’t make that face: a report on anthropomorphizing and interface. *AAAI’98* 173-179.
47. Yarkoni, T. Personality in 100,000 words: A large-scale analysis of personality and word usage among bloggers. *J. Res. in Personality*, 2010, 44(3): 363–373.
48. Yu, R., Liu, Y., & Yang, M. Does interviewer personality matter for survey outcomes? *Proc. 64th Annual Conference on World Association for Public Opinion Research*, 2011.
49. Ziegler, M., MacCann, C., & Roberts, R. (editors). *New perspectives on faking in personality assessment*. 2010, Oxford University Press.