Ghosts in the Machine: Personalities for Socially Adroit Software Agents

D. Christopher Dryer
International Business Machines Corporation, Almaden Research Center
650 Harry Road
San Jose, CA 95120 USA
+1 408 927 1912
dryer@almaden.ibm.com

Abstract

Recent research indicates that people respond socially to computers and perceive them as having personalities. Software agents especially are human-machine interaction artifacts that embody those qualities most likely to elicit social responses: contingent behavior, fulfilling a social role, and using language. People may perceive agents as one or many discreet social entities, each with a personality that will shape people's interactions with the machine. An industrial team's research on and experience with the design of software agent personalities is described, in the context of related social science and computer science research. The conclusions from this experience are summarized as guidelines for future agent developers.

Keywords
Agents; Design rationale; Guides; Personality; Social Interface; Theory; Wizards

About the project

You have probably heard that we treat computers like people. That is, we are polite to them, we perceive them as teammates, we like it when they compliment us, and so on. In numerous studies (e.g., Dryer et al. 1993a; Dryer et al. 1993b; Nass et al. 1994; for summary, see Reeves and Nass 1996), researchers have demonstrated that people use natural and social rules to understand the behavior of technologies. Indeed, we even perceive computers as having a personality.

What you may not have heard is that we may perceive a computer as having many personalities. A single machine does not necessarily have a single personality. Recent studies (Dryer et al. 1993b; Moon and Nass forthcoming; Nass et al. 1995; Reeves and Nass 1996) suggest that the design of user interfaces may determine when people perceive one or more social personalities. Different social partners are perceived as having distinct personalities, and different personalities are evidence for distinct partners.

Unfortunately, the various personalities in any given machine are likely there by chance rather than by design. Ten different error messages may have been written by ten different authors, each with a different style. The result is not the absence of personality but instead a confusing or vague personality at best and a frustratingly schizophrenic personality at worst. The research (Reeves and Nass 1996) suggests that people will see a personality in the machine regardless of whether one was intended.

These findings open the door to some new and exciting opportunities for human-machine interaction designers. Specifically, how can software agents be designed so that their personalities support satisfying social interactions? This paper concerns recent work on designing software agents, especially "WarpGuide." WarpGuide is a task mentor in IBM's operating system, OS/2 Warp 4 (Boehm et al. forthcoming; Selker 1994). To create a personality for WarpGuide, an interdisciplinary team used the theory and research from the fields of intelligent agents and social psychology to inform our design decisions. This is an unusual exercise in applying psychological research to software agent design and, as far as we know, this is the first time the issue of "one or many?" personalities has been considered.

Theoretical Background

Personality is a collection of individual differences, dispositions, and characteristic adjectives with some consistency across situations and time. Social behavior is complex and important. Successful social behavior negotiates the goals, beliefs, and emotions of multiple social partners. These negotiations occur in most of the domains of human life: collaborative work, recreation, family life, and so on. People automatically and unconsciously use personality as a tool to organize otherwise overwhelming information about social partners. Personality helps people make predictions about what Person A will do and how that is different from what Person B will do. Without personality and the ability to use personality to organize these predictions, social behavior would break down.

Interestingly, human personalities are infinite. Mapping the space of the universe of possible personalities may seem like an impossible task. Indeed, there are thousands of non-synonymous trait adjectives (like shy, adventurous, obsessive, domineering, feminine, anxious, etc.).

As it happens, however, the task is a bit simpler than it seems. It turns out that, while there are thousands of different facets to personalities, there only a few things that really matter at a more abstract level. The five important factors (for review, see McCrae and John 1992) can be described as: (1) Agreeable (cooperative to competitive); (2) Extroverted (outgoing to withdrawn); (3) Neurotic (anxious to calm); (4) Conscientious (organized to lax); (5) Open (curious to closed-minded). Certainly, there are other things to know about a partner, but nearly all of them covary with one of these five things or some combination of them.

These factors result from empirical investigations into which sets of individual differences (among the universe of all individual differences) covary. As it turns out, all of the theoretically derived personality schemes (like the Myers-Briggs [Croom et al. 1989]) that have been tested can be shown to be simply rotations of some or all of the Big Five dimensions.

User Interface Agents

> Direct manipulation has its place, and in many respects is part of the joys of life: sports, food, sex, and for some, driving. But wouldn't you really prefer to run your home and office life with a gaggle of well-trained butlers [...] and, on some occasions, cooks, gardeners, and chauffeurs when there were too many guests, weeds, or cars on the road? (Negroponte 1990)

Software agents have begun to have widespread application. In this new territory, researchers, designers, and developers are finding new ways to help people get the most from their machines by delegating some of the work to agents. In particular, ongoing work in the area of intelligent agents concerns the application of artificial intelligence (AI) technologies to the problems of human-machine interactions (Miller, Sullivan, Tyler 1991). As future software agents become more widespread, they will need to become more intelligent in the social domain.

Software agents, such as WarpGuide, have three important features. (1) They can use full sentence text, in addition to the typical user interface (UI) forms of communication, like menus, controls, and icons. Full sentence text is more natural, especially for people with a more verbal than nonverbal cognitive style (Horn and Cattell 1966). (2) They can embody task knowledge as well as use AI to reason about when and how to engage a person in interaction. This gives them a compelling kind of contingent behavior. (3) They can autonomously perform actions on a person's behalf, much like the social roles of butlers, cooks, gardeners, and chauffeurs. These features are important because natural language (text), contingent behavior (intelligence), and social role (autonomous assistance) are the three fundamental predictors that people will respond socially to a human-machine interaction artifact (Reeves and Nass 1996).

Social interface theory is built on a program of studies demonstrating that people respond socially to machines (Reeves and Nass 1996). These studies suggest that technology equals real life. Whenever possible, people automatically and subconsciously leverage what they know about their natural and social experiences to help them with their technological experiences. Because we are first and foremost human, we are inclined to treat everything as social and natural.

A UI, however, is not necessarily like a single physical object or single social partner. Rather, a user interface is like a proscenium to a stage populated by people, places, and things (Laurel 1986). The stage may contain a single actor or an entire troupe. In graphical user interfaces, it is easy to see how the icons might be, psychologically, distinct physical objects. In the same way, autonomous interface personalities might be distinct social partners.

As an example, researchers have demonstrated that social entities and voices are matched one to one. In a laboratory study, two computers with one voice were perceived as a single social partner. One computer with two voices was perceived as two social entities (Dryer et al. 1993b). In other words, voice is one clue that people use to integrate and distinguish collections of behaviors into discrete social entities.

Personality is another clue that people use to organize behaviors. In interpersonal interactions, the personality of the participants determines how enjoyable and productive the interaction is (Dryer, 1993; Dryer and Horowitz, 1997;

Horowitz, Dryer, and Krasnoperova 1997]. In people's interactions with technology, they automatically use personality to organize behavior (Dryer et al. 1993a; Nass, Reeves, and Dryer 1996). People's own personality determines how they respond to the technology (Detenber and Dryer, 1995). The perceived personality of the technology determines how much people enjoy the interaction (Moon and Nass forthcoming; Nass et al. 1995). Moreover, the fundamental personality dimensions of all interactions, both interpersonal and human-machine, are the same (Dryer 1993; Dryer and Horowitz 1997; Dryer, Jacob, and Shoham-Saloman, 1994; Dryer et al. 1993a; Reeves and Nass 1996; Wiggins 1979). Researchers have intentionally created social entities by designing user interfaces (Nass et al. 1995) and agents (Bates 1994; Hayes-Roth, Brownston, and Sincoff, 1995; Maes, 1995; Oren et al. 1990) to have specific emotions, states, roles, and personalities.

Laboratory Tests

To investigate how we might design personalities for socially adroit software agents, we conducted a series of laboratory studies in which thirty-one participants encountered presentations of thirty-seven animated characters. We assessed the participants' personalities, their perceptions of the characters along the five personality factors, and their liking of the characters. We were interested in the aspects of the characters personalities that made them likable. From this research, we came the following conclusions:

(1) "Positive" personalities are liked. Each of the five dimensions has a positive, or socially desirable pole, and a negative pole. In general, personalties that are described at the positive end of the poles are better like than the opposite.

(2) Strong personalities are liked. Personalities are better liked when they are extreme along at least one of the five dimensions. In fact, it may be better to extreme on the socially undesirable end of a dimension than in the middle. Rarely, personalities are so extreme that they appear unnatural and are disliked. In practice, however, it is difficult to create a personality that is so extreme it is disliked.

(3) Personalities with a foible are liked. Personalities that are entirely positive are disliked. Well-liked characters tend to have a negative personality attribute along one dimension.

(4) Personalities that are similar to a person's own are liked. In general, people like partners who have personalities that are similar to their own.

(5) Personalities that make up for something you do not have are liked. There are special situations in which people prefer partners who are complementary to them rather than similar to them. For example, when a task requires one person to take charge, a dominant person will prefer to work with a less dominant partner.

(6) Consistent (or meaningfully changing) personalities are liked. Interesting personalities tend to be expressed consistently across time and situations. As an exception, personalities can also be interesting when they change. What is important is that the change is clear and meaningful. When a dominant partner follows your lead rather than challenging your control, it has meaning because it communicates something, such as a respect for your competence. Personality "inconsistencies" work when they turn out to be clearly "consistent" after all. Moreover, personality consistency does not require that a person do the same thing all the time. That would be unnatural. Instead, interesting personalities get expressed in a myriad of various, specific, unique, and even surprising ways.

(7) Specifically expressed personalities are liked. The five factor model captures the structure of our perceptions of personality, but not how personality is naturally expressed. For example, saying that someone is "open" is nearly meaningless; the experience of personality occurs at a more basic level. For example, an open person might be someone learning the violin, fond of wine tasting, and an avid sky diver. Part of what is intriguing about personalities is how one personality can be expressed in different ways across different situations.

The Experience with WarpGuide

WarpGuide uses software agent technology to guide people through certain system tasks, helping to prevent them from making errors. WarpGuide communicates in full-sentence text and responds to voice commands (just like the other components of OS/2 Warp 4). Because of WarpGuide's social role, use of text, and perceived intelligence, we knew it would be important to consider people's social responses. Specifically, we addressed the questions: (1) would WarpGuide's function be designed best as the behavior of a single social partner or many; (2) what personality would be best for WarpGuide; and (3) how was that personality best expressed?

We first considered the argument for multiple social personalities. People perceive whatever is present as the social partner (Nass and Sundar 1996). WarpGuide's behavior might make it salient as a social partner, distinct from other behaviors. Moreover, this perception might be advantageous. Separate social personalities may be used in a "good cop/bad cop" routine. That is, one partner could handle all emotionally negative aspects of an interaction leaving interactions with another partner emotionally positive (Reeves and Nass 1996).

We therefore decided that WarpGuide needed to be distinct from the "system homunculus." The idea was that, psychologically, people might have troubled relationships with their primary partner, the system. As a partner, a system may be hard to get along with sometimes. People's relationship with WarpGuide would be different. WarpGuide would be a distinct social partner, whose only job was helping with tasks. It would be like a voice over the shoulder, guiding people through their interactions with the system. This way, WarpGuide would be "blameless" for whatever else the system did.

The next decision concerned whether WarpGuide, as a distinct social partner, would itself be a single partner or multiple personalities. Here, we had to consider two competing arguments. On the one hand, separate social personalities can have the advantage of being "specialists." By virtue of having a narrow range of expertise, people both perceive specialists as being better than "generalists" for that domain and tend to have more realistic expectations for their behavior (regardless of whether these assumptions are true) (Reeves and Nass 1996). On the other hand, people like simple. Managing multiple social relationships can get complicated, turning a potentially simple one-on-one interaction into a group relationship (Reeves and Nass 1996).

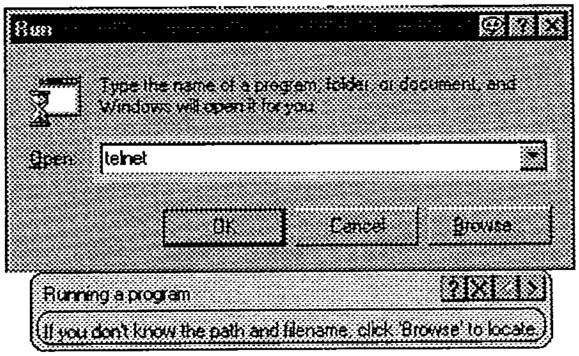
In our case, WarpGuide would guide people through different tasks. Certainly, technologists talk about one agent for this task and another for that. Was WarpGuide best thought of as a collection of entities, each being an expert at a single task?

Before we answered that question, another question occurred to us about WarpGuide's behavior. WarpGuide would behave in different ways depending on the task. We had both guide technology (COACH) and wizard technology (SmartGuide) available to us. Because each technology has its advantages for different tasks (Dryer 1997; Wilson 1995), we wanted WarpGuide to have both behaviors. Did this difference in behavior mean that we had to create WarpGuide as two separate social personalities?

We decided that what people really wanted was a single social partner that they could turn to for help. WarpGuide would be like that expert down the hall, someone you turn to whenever you have a problem with your computer. To simplify the interaction, people would have a single, social relationship with one task mentor. WarpGuide would have expertise across a set of tasks but would be an expert at mentoring only. Moreover, people wouldn't care what underlying technology (guide or wizard) WarpGuide was using for guidance; they would just want the best guidance for the particular task. So, WarpGuide was conceived as a single personality with two roles: guide and wizard.

To complete our work, we needed to integrate WarpGuide's roles and behaviors into a single social partner. For OS/2 Warp 4, we did not have the option of giving WarpGuide a single voice, at least not one that could be heard. Instead, we needed to find other ways to create a single, consistent personality. Starting with the research on personality, we considered what kind of personality would be best for WarpGuide. In general, people prefer friendly personalities. With respect to control, the appropriate personality really depends on the kind of interaction (Reeves and Nass, 1996) For a mentoring interaction, we decided WarpGuide needed to be authoritative but not so controlling that it was perceived as intrusive. Therefore, we defined the personality as intelligent, friendly, and unobtrusive. We then explored ways to create this personality.

To communicate a personality, an entity's behavior needs to be consistent (Reeves and Nass, 1996). We therefore endeavored to design WarpGuide's behavior to be consistent in all its functions. As one example, WarpGuide offers a consistent starting point for many different tasks. It presents a collection of "guidance objects" organized in a single location. When a guidance object is opened, WarpGuide brings up the appropriate task interface along with the accompanying assistance. WarpGuide also provides a way to navigate through tasks, from beginning to end, even across multiple task interfaces that normally may be scattered throughout the user interface. WarpGuide works the same way across both roles and across a range of tasks, unifying these roles and behaviors into a single personality.



Type the name of a program, folder, or document, and Windows will open it for you.

Open: telnet

Running a program

If you don't know the path and filename, click 'Browse' to locate.

Figure 1. WarpGuide appearance.

Visual appearance is another clue to personality (Nass, Reeves, and Dryer 1996; Reeves and Nass 1996). WarpGuide is represented graphically in the UI by a set of objects, windows, icons, and annotations. All these elements are unified by common visual themes: in particular, a characteristic cue card form, an annotated dialog, certain colors, and iconic "eyes" (suggesting its social role). The colors and forms are designed to be "soft" and yet authoritative, representing friendliness and intelligence. Specifically, we used "friendlier" rounder shapes, such as the rounded rather than angular corners of the cue card, and we went after a "friendly" look in the design of the iconic eyes, attempting to use their form to represent the right personality. All the elements are unified to some extent by the color yellow. Because WarpGuide's form is a set of visual elements that are variations on a single theme, that form helps to express a single personality.

The next step in building WarpGuide's personality was to consider its message content. Research has shown that text alone is a powerful clue to personality (Moon and Nass forthcoming; Nass et al. 1996). We therefore fashioned a style of speech around strict guidelines. We developed templates for sentence structure and wrote each text article to fit the templates. In addition, multiple team members reviewed each article, to insure that (1) the content was intelligent; (2) the tone and phrasing were friendly; and (3) the structure was consistent, concise, and unobtrusive. As an example, an article for the finding a file dialog is "Type as much of the name as you know, and use an asterisk (*) for the rest." A version of the sort "You can type only as much as you know, and use an asterisk (*) for the rest" makes the software agent seem less authoritative. A version of the sort "If you don't know the whole name, type what you know and use a asterisk for what you don't know" makes the software agent seem less friendly. Of all the work, the text was probably the most important to establish WarpGuide's personality.

Finally, although we used behavior, images, and text phrasing deliberately to construct a personality, we were also careful to avoid anthropomorphism and overuse of social metaphors. Our goal was to make the most of social responses that people would have, not artificially create a social relationship. In particular, we avoided a full face or figure and avoided using "I" in the text. WarpGuide was not meant to be "cute," nor was it meant to portray itself as "human." Instead, our intent was that it would behave socially in ways that make sense given its context.

Generalizations: Tips for designing personalities

As designers become more aware of the impact social responses have in human-machine interactions, they will face some of these same issues. Our experience can be summarized in these guidelines:

1. Language use, fulfilling a social role (like mentoring), and perceived intelligence are likely to encourage social responses.

2. Multiple personalities can exist in a user interface. If you are likely to have a social response, consider designing one or more discrete personalities to manage that relationship.

3. Well-liked personalities tend to be socially desirable, to be strong, to include a foible, to be similar or complementary, to be consistent or to change in meaningful ways, and to be expressed at a natural, basic level.

4. If people are likely to be frustrated, consider designing a personality they can blame and another that can help them.

5. People like simple; integrate as much behavior as possible into a single personality. Technological distinctions (like guide versus wizard) may not be psychologically relevant.

6. Create a social partner in a way that makes sense for the context. The goal is not to duplicate mindlessly the natural and social world in cyberspace; faces and the pronoun "I" may not be appropriate.

Acknowledgments

This work was the product of the WarpGuide team, including: Ron Barber, Roger Didio, Steve Idhe, Bob Kamper, Bob Kelley, Marion Lindsey, Craig Moser, Paul Rogers, Ted Selker, John Tyler, Les Wilson, and Julie Wright. IBM and OS/2 Warp are trademarks of the IBM Corporation.

References

Attkinson, B., Brady, S., Gilbert, D., Levine, D., O'Connor, P., Osisek, D., Spagna, S., Wilson, L. 1995. IBM Intelligent Agents. *UNICOM Seminar Proceedings.*

Joseph Bates. 1994. The Role of Emotion in Believable Characters. *Communications of the ACM, 37,* 7.

Boehm, L., Efruss, R., Fraley, B., Magid, P., Robinson, A.M., Schneider, I., and Wilson, L. Forthcoming. The New Workplace Shell in OS/2 Warp 4, *Personal Systems.*

Croom, W.C., Wallace, J.M., Schuerger, J.M. 1989. Jungian types from Cattellian variables. *Multivariate Experimental Clinical Research, 9,* 1, 35-40.

Detenber, B. and Dryer, D.C. 1995. Eye of the beholder: Individual differences in responses to mediated images. Presented at the annual meeting of the International Communication Association, Sante Fe, NM.

Dryer, D.C. 1993. Interpersonal goals and satisfaction with interactions. Doctoral dissertation, Stanford, CA: Stanford University.

Dryer, D.C. 1997. Wizards, guides, and beyond: A rational and empirical method for selecting optimal intelligent user interface agents. *Proceedings of the Intelligent User Interface conference*, Orlando, FL.

Dryer, D.C., and Horowitz, L.M. 1997. When do opposites attract? Interpersonal complementarity versus similarity. *Journal of Personality and Social Psychology, 72,* 3, 592-603.

Dryer, D.C., Jacob, T., and Shoham-Saloman, V. 1994. The communication style typology of families. Presented at the annual meeting of the American Psychological Association, Los Angeles, CA.

Dryer, D.C., Nass, C., Steuer, J.S., and Henriksen, L. 1993. Interpersonal responses to computers programmed to function in social roles. Presented at the annual meeting of the Western Psychological Association, Phoenix, AZ.

Dryer, D.C., Steuer, J., Henriksen, L., Tarber, E., and Reeder, H. 1993. Computers are social actors II: Voice is the mirror of the soul. Presented at the annual meeting of the International Communication Association, Washington, D.C.

Hayes-Roth, B., Brownston, L. & Sincoff, E. 1995. Directed Improvisation by Computer Characters. Knowledge Systems Laboratory, Stanford University, Technical Report KSL-95-04.

Horn, J.L., and Cattell, R.B. 1966. Refinement and test of the theory of fluid and crystallized ability intelligences. *Journal of Educational Psychology, 57*, 253-270.

Horowitz, L.M., Dryer, D.C., and Krasnoperova, E.N. 1997. The circumplex structure of interpersonal problems. In R. Plutchik & H.R. Conte (eds.), *Circumplex Models of Personality and Emotions.* American Psychological Association: Washington, D.C..

Laurel, B. 1986. Interfaces as mimesis. In D.A. Norman and S.W. Draper (eds.) *User Centered Design: New Perspectives on Human-Computer Interface.* Lawrence Erlbaum Associates: Hillsdale, NJ.

Maes, P. 1995. Artificial life meets entertainment: Lifelike autonomous agents, *Communications of the ACM, 38,* 11, 108-114.

McCrae, R.R., and John, O.P. 1992. An introduction to the five-factor model and its applications, *Journal of Personality, 60,* 175-215.

Miller, J.R., Sullivan, J.W., and Tyler, S.W. 1991. Introduction. In J.W. Sullivan and S.W. Tyler (eds.) *Intelligent User Interfaces.* Addison Wesley: New York, NY.

Moon, Y., and Nass, C. How "real" are computer personalities? *Communication Research,* in press.

Nass, C., Moon, Y., Fogg, B.J., Reeves, B., and Dryer, D.C. 1995. Can computer personalities be human personalities? *International Journal of Human-Computer Studies, 43*, 223-239.

Nass, C., Reeves, B., and Dryer, D.C. 1996. Adults and novel cartoon characters. In B. Reeves & C. Nass, *The Media Equation.* Cambridge University Press: Cambridge, 81-82.

Nass, C., Steuer, J.S., Henriksen, L., and Dryer, D.C. 1994. Machines and social attributions: Performance assessments of computers subsequent to "self-" or "other-" evaluations. *International Journal of Human-Computer Studies, 40*, 543-559.

Nass, C. and Sundar, S.S. 1996. Source orientation. In B. Reeves & C. Nass, *The Media Equation.* Cambridge University Press: Cambridge, 181-189.

Negroponte, N. 1990. Hospital corners. In B. Laurel (ed.) *The Art of Human-Computer Interface Design.* Addison Wesley: Reading, MA.

Oren, T., Salomon, G., Kreitman, K., and Don, A. 1990. Guides: Characterizing the interface. In B. Laurel (ed.) *The Art of Human-Computer Interface Design.* Addison Wesley: Reading, MA.

Reeves, B. and Nass, C. 1996. *The Media Equation.* Cambridge University Press: Cambridge.

Selker, T. 1994. Coach: A teaching agent that learns. *Communications of* the ACM, 37, 7, 92-99.

Wiggins, J.S. 1979. A taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*, 395-412.

Wilson, L. 1995. Intelligent agents: A primer. *Personal Systems, September/October*, 47-49.