Association for Information Systems
AIS Electronic Library (AISeL)

SIGHCI 2014 Proceedings

Special Interest Group on Human-Computer Interaction

2014

Inferring Capabilities of Intelligent Agents

Bart Knijnenburg
UC Irvine, bart.k@uci.edu

Martijn C. Willemsen
Eindhoven University of Technology, M.C.Willemsen@tue.nl

Inferring Capabilities of Intelligent Agents from their External Traits

Bart P. Knijnenburg
University of California, Irvine
bart.k@uci.edu

Martijn C. Willemsen
Eindhoven University of Technology
m.c.willemsen@tue.nl

ABSTRACT

We investigate the usability of human-like agent-based interfaces. In an experiment we manipulate the capabilities and the "human-likeness" of a travel advisory agent. We show that users of the more human-like agent form an anthropomorphic use image of the system: they act as if the system is human, and try to exploit typical human-like capabilities. Unfortunately, this severely reduces the usability of the agent that looks human but lacks human-like capabilities (overestimation effect). We also show that the use image users form of agent-based systems is inherently integrated (as opposed to the compositional use image they form of conventional GUIs): cues provided by the system do not instill user responses in a one-to-one manner, but are instead integrated into a single use image. Consequently, users try to exploit capabilities that were not signaled by the system to begin with, thereby further exacerbating the overestimation effect.

Keywords

Agent-based interaction, anthropomorphism, usability, feedforward and feedback, use image.

INTRODUCTION

Agent-based interaction, in which the user interacts with a virtual entity using natural language, has been a topic of HCI research for several decades (Qui and Benbasat 2009; Walker et al. 1994; Quintanar et al. 1987; Nickerson 1976), and has gained renewed attention with the rise of smartphone agents like Siri, Cortana and Google Now. Because agent-based interaction is finer grained and richer than interaction with conventional Graphical User Interfaces (GUIs), it should better suit the increasingly complex tasks we perform with computers (Laurel 1990). People also find agent-based interaction more enjoyable and more natural (Kang et al. 2012; Hess et al. 2005). At the same time, though, some agents remind us of "Clippy": they seem unable to live up to their promises (Nowak 2006; Dehn and van Mulken 2000). In this paper we address the usability of agent-based interaction, and identify a cognitive principle that makes agent-based interaction different from traditional GUIs.

RELATED WORK AND THEORY DEVELOPMENT

To explain why some systems are more usable than others, Norman (1986) argues that there are two gulfs between the user and the system: the gulf of execution (the user has to discover how to manipulate the system to accomplish a task) and the gulf of evaluation (the user has to interpret whether output of the system is in line with their goal). Users bridge these gulfs by forming a "use image", a mental representation of the way the system works that helps them to infer which interface actions fulfill their goal, and what the output of the system means. According to Norman, the formation of an adequate use image is greatly facilitated by providing appropriate feedback (e.g. responses to actions) and feedforward (e.g. labels on buttons).

The Layered Protocol Theory (LPT) operationalizes Norman's use image theory (Taylor 1988). It decomposes user-system interaction into a set of layers, each breaking users' intentions down into smaller components. Brinkman (2003) argues that this compositional character is reflected in the users' use image: the *compositional* use image is the sum of the use images of its widgets (e.g. levers, buttons, text fields, scrollbars).

Agent-Based Use Images

Many usability researchers and designers have assumed the compositionality of the use image. Most usability evaluation techniques evaluate the different parts of an interface separately; the effectiveness of these techniques thus depends on the legitimacy of the compositional use image. Compositionality seems to hold for "real life" interfaces (e.g., doors, phones) as well as conventional GUIs. However, agent-based interfaces typically lack the common levers, buttons, text fields and scrollbars. So how do users form a use image of agent-based interfaces?

Cook and Salvendy (1989) note that users infer the use image of an agent-based system from the way it "looks" and "talks" (feedforward) and the apparent intelligence of its responses (feedback), just like they would do when interacting with other human beings. In fact, Laurel (1990) argues that users attribute common human *intelligence* to systems that provide human-like appearance and capability cues. Indeed, studies show that users of systems with a cartoon character that "talks" in full sentences and personifies itself believe that it shows some form of human intelligence, while users do not show similar beliefs when using a system without such a cartoon character that talks "computerese" (De Laere et al. 1998; Quintanar et al. 1987). We thus argue that:

The more human-like the system looks (appearance cues) and the more capabilities it displays (capability cues), the more intelligent users believe the system to be.

Note that the *actual* capabilities of the system might not necessarily co-occur with capability *cues*; the system might exhibit specific linguistic capabilities (e.g. using the word "here" to refer to the current location) without actually being able to understand them in the user dialog (e.g. it may not be able to infer the current location when the user uses the word "here"). In effect, cues of human-like appearance and capabilities can underplay or overplay the agent's actual capabilities.

What psychological mechanisms could underlie the use image of believed intelligence? Thompson (1980) found that users of a natural language-based system showed a tendency to *anthropomorphize* the behavior of the system, and this tendency also increases with personalization, conversational tone, affective responses and diversified wording (Quintanar et al. 1987; De Laere et al. 1998). Not only agents are subject to anthropomorphism: users of any computer system occasionally engage in negative anthropomorphism (e.g. shouting; Chin, et al. 2005), and adhere to social principles (e.g. politeness effect; Reeves and Nass, 1996). Bradshaw (1997) argues that when a system's behavior is too complex to understand, users are inclined to take the "intentional stance" (Dennett 1987): they attribute intentional behavior to systems as a convenient shortcut towards explaining complicated behavioral patterns (i.e. the system "*wants* me to do X", or "*doesn't like* it when I do Y"). This then also leads them to adhere to human social principles. The intentional stance holds for any system, but agents instill stronger anthropomorphic reactions (Nowak 2006). Therefore:

In agent-based systems, the intentional stance is at the heart of the use image construction. The use image is an anthropomorphic construct, instilled by human-like cues.

As the use image is a mental construct, one cannot observe directly whether or not it is anthropomorphic. However, reactions to the use image can provide evidence of its nature: if the use image is anthropomorphic, users will interact with the agent in a way that is in accordance with human-human interaction. Examples of "Human-like responses" are the use of long and grammatically correct sentences. Indeed existing research has found that the use of a human-like avatar and personalized feedback (human-like cues that may lead to an anthropomorphic use image) leads users to be more verbose in their responses (Brennan 1991; Rosé and Torrey 2005; Walker et al. 1994; Richards and Underwood 1984). In sum:

Since the use image of an agent is anthropomorphic, users will act in a more human-like way towards a system they believe to be more intelligent.

Moreover, if the system looks and behaves human, then users will believe it has typical human capabilities, and will try to exploit these capabilities. An important category of human capabilities is the linguistic capability of implicit *reference* to the context of the conversation (Levinson 1983; Halliday and Hasan 1976). Computers are notoriously bad at understanding such references (Winograd 1972; Dey 2001; Scheutz et al. 2011), but users may believe that human-like systems, like human beings, can resolve them. Specifically, they may believe that agents can understand references to a mentioned location, (e.g. "here", "there"), time (e.g. "now", "then"), or object ("that trip", "that ticket"). In sum, we argue that:

Users will assume that systems they believe to be more intelligent have more advanced linguistic capabilities, and they will try to exploit these capabilities.

An Integrated Use Image?

If agent-based interfaces were like traditional GUIs, their use image would be compositional. There would be a one-to-one mapping where each cue would instill its own use image and induce a corresponding response. Brennan (1991) found support for such a one-to-one mapping in both human-human and natural language-based human-computer interaction. Participants in her experiments showed *syntactic entrainment*; a direct reflection of the conversation partner's responses. According to these findings, one could evoke a certain behavior in the users' response by expressing the same behavior in the agent.

However, the intentional stance (Dennett 1987) should allow users to create an *integrated* use image based on the behavior of the system *as a whole*. If the system is sufficiently human-like, it will be attributed intentional behavior, and this attribution is based on the "human-likeness" of the agent as a whole, not on a specific part of its behavior. The fact that the "system" is "human" provides them *instantaneously* and *effortlessly* with a detailed use image of what it can do and how to interact with it: if the system looks and behaves human, the use image simply dictates that the system can and cannot do whatever humans can and cannot do. In the words of Laurel (1990, pp. 358-359): "[An agent] makes optimal use of our ability to make accurate inferences about how a character is likely to think, decide and act on the basis of its external traits. This marvelous cognitive shorthand is what makes plays and movies work […] With interface agents, users can employ the same shorthand—with the same likelihood of success—to predict, and therefore control the actions of their agents." However, if users *integrate* the system cues into a single use image of believed intelligence, this creates a much less straightforward relation between system cues and user responses:

All cues about the intelligence of the system will be integrated into a single use image and instill a series of possibly unrelated responses.

The integrated use image has both positive and negative consequences for agents' usability (Dehn and van Mulken 2000; Qiu and Benbasat, 2009). The integrated use image makes agents especially suitable for complex tasks. Conventional GUIs require additional widgets for each additional function, which makes it impossible to create a

really usable GUI for a complex system. Agent-based interfaces, on the other hand, have an integrated use image that instantly provide users with a heuristic to determine what they can and cannot do, and how to access the functionality. However, if the agent looks more capable than it actually is, users will *overestimate* the system's capabilities, which will result in confusion and reduced usability (Brennan 1991; Forlizzi et al. 2007; Walker et al. 1994).

Good usability arises when the user tries to use only those capabilities that the system actually provides. This means that the use image has to match the actual system capabilities (Norman 1986). If the use image of an agent were compositional, it would be fairly easy to "manage" this use image: the system could simply provide a matching cue for each capability. However, an integrated use image is much harder to manage, because there is more than just a one-to-one relation between cues and responses. In effect, even human-like *appearance* cues may instill *capability*-exploiting responses: merely "looking human" may be enough to make users believe that the system has certain human-like capabilities (even if these are not actually present).

In sum, the presumed integrated use image is responsible for both the greatest advantage but at the same time the most significant drawback of agent-based interaction: due to our natural tendency to use anthropomorphism, it is very easy to *instantly* create a complex, integrated use image from which users can effortlessly infer a myriad of complex functions to perform with the system, along with possible ways to exploit them. However, since these functions are not directly coupled to a specific underlying cue, an overestimation effect can easily occur, and it will be rather difficult to tweak this use image such that it perfectly matches the actual system capabilities.

HYPOTHESIS DEVELOPMENT

The goal of our experiment is to empirically demonstrate that the use image of agents is more likely to be integrated rather than compositional. We also want to test if an agent-based use image indeed instills human-like and capability-exploiting responses. Finally, we want to evaluate the effect of these responses on the usability of the interaction (i.e. the overestimation effect).

In our experiment, we independently vary the system's feedforward cues and its actual capabilities. There are three levels of **cues**: "computer-like cues" (agent looks and talks like a computer), "human-like appearance cues" (agent looks and talks like a human being), and "human-like appearance and capability cues" (agent additionally uses references in its sentences, which signals its capability to understand such references). There are two levels of **capabilities**: "low capabilities" (system can only process simple, complete requests) and "high capabilities" (system can process complex requests with implicit references, like a human being). We argue that a system with high capabilities should generally be easier to use:

H1. Usability in the "high capabilities" condition should be higher than in the "low capabilities" condition.

Within the low capabilities conditions overestimation can occur, when the system appears human-like and displays human-like cues:

H2. Within the "low capabilities" conditions, the "human-like appearance cues" and "human-like appearance and capability cues" conditions should lead to lower usability than the "computer-like cues" condition.

Moreover, the existence of an anthropomorphic use image predicts the presence of more human-like and capability-exploiting responses when human-like cues are provided:

H3. Within the "high capabilities" conditions, the "human-like appearance cues" and the "human-like appearance and capability cues" conditions lead to more human-like and capability-exploiting responses than in the "computer-like cues" condition.

Finally, if the agent-based use image is compositional, appearance cues cannot evoke capability-exploiting responses, and users will try to exploit human-like capabilities in the "human-like appearance and capability cues" condition *only*. But if the use image is integrated, a human-like *appearance* cue given by the agent can evoke *capability*-exploiting responses, and users will try to exploit human-like capabilities in both the "human-like appearance cues" *and* the "human-like appearance and capability cues" conditions. In other words, if the following hypothesis is upheld, this would rule out the compositional use image would predict, and provide evidence for an integrated use image:

H4. Within the "high capabilities" conditions, users exhibit more capability-exploiting responses than in the "computer-like cues" condition even when the system does not give human-like capability cues (i.e. even in the "human-like appearance cues" condition).

Experimental setup

For the experiment we created an online agent that gives travel info for the Dutch Railways. 92 university students from all over The Netherlands (35 male; age M=21.8, SD=3.55) took part in the experiment. For additional power to test H3 and H4, 59 participants were randomly assigned to the "high capabilities" conditions and only 33 to the "low capabilities" conditions. Users performed four predefined tasks (e.g. "You are in Eindhoven and you want to go to Tilburg. Find out if you have to switch trains anywhere.") by typing requests to the system. A Wizard of Oz technique was used to provide the answers: users were ostensibly interacting with a real system, but were actually talking to the experimenter, who read inputs and provided responses using a strict protocol.

We measured personal references, number words per request, and grammatical correctness of requests as **human-like responses**. These behaviors typically occur in interaction between two humans, but not when interacting

with a computer (Rosé et al. 2005; Shechtman and Horowitz 2003; Brennan, 1991). We measured explicit and implicit references to times, places, and earlier questions, and asking multiple questions at once, as **capability-exploiting responses**. These behaviors occur when a participant assume that the agent understands the context of the conversation, like a human would. Finally, we measured **usability** by the number of requests and amount of time participants needed per task (efficiency), the difference in time per task between the first and last task (learnability), the "overall reactions to the software" section of the QUIS (Chin et al. 1988), and occurrences of users discontinuing the experiment (effectiveness).

RESULTS

We first confirm that the system with high capabilities is actually more usable than the system with low capabilities (H1) Users needed fewer requests per task (2.38 vs. 3.32, $p < .005$), and were more satisfied (31.47 vs. 24.07, $p < .001$) in the high-capabilities than the low-capabilities conditions, indicating that the former was indeed more usable than the latter. The average time needed to perform the tasks was actually higher in the high capabilities conditions than the low capabilities conditions (211s vs. 185s, not significant).

H2 suggests that for low capability systems, users in the "human-like appearance cues" and "human-like appearance and capability cues" conditions overestimate the capabilities of the system, resulting in lower usability than the "computer-like cues" condition. Strong evidence for overestimation was found in terms of system effectiveness: 5 of the 23 participants interacting with a system with low capabilities but human-like cues (and none for computer-like cues) prematurely quit the experiment. Additional evidence of overestimation was found in terms of learnability. Within the low-capabilities condition, users of the computer-like interface showed a higher time decrease from task 1 to task 4 (–108.56s) than users of the human-like systems (–40.12s and –25.89s, $p < .05$).

H3 suggests that human-like responses in the "high capabilities" conditions increase from computer-like cues, to human-like appearance cues, to human-like appearance and capability cues. Figure 1 shows that first-person references ($\beta = 1.16$, $p < .001$), words per chat request ($\beta = 2.50$, $p < .005$) and grammatical correctness ($\beta = 6.47$, $p < .005$) indeed increased with cue level. This is evidence for the existence of an anthropomorphic use image, as several human-like responses were significantly higher when the system had more human-like cues.

H3 also suggests that the occurrence of capability-exploiting responses in the "high capabilities" conditions increases with cue level. A sum measure of the five capability-exploiting responses (see Measures) was taken for each task. Figure 1 (rightmost panel) shows that the number of capability-exploiting responses in the human-like conditions is significantly higher than in the computer-like cues condition ($\beta = 0.288$, $p < .05$). These results provide further evidence for the existence of an anthropomorphic use image, as the total number of capability-exploiting responses was significantly higher when the system had more human-like cues.

Finally, if users have an integrated use image (H4), they should show capability-exploiting responses even when the system does not give human-like capability cues (i.e. when it gives human-like appearance cues only). Figure 1 (bottom panel) shows that capability-exploiting responses are indeed higher in the "human-like appearance cues" systems than in the "computer-like cues" condition (a planned contrast between "computer-like cues" and the other two conditions is significant at $p < .05$). This rules out a compositional use image, since it would require that only the capability cues condition can induce capability-exploiting responses. In fact, capability-exploiting responses in the "human-like appearance cues" condition are not significantly different from the "human-like appearance and capability cues" condition.

CONCLUSION AND MANAGERIAL IMPLICATIONS

Siri, Cortana and Google Now show that agents definitely have potential. Still, managers have to be very careful introducing a human-like agent in their systems. Human-like agents are a *metaphor*; its cues are effortlessly *integrated* into a single use image. This use image, though, instills a set of responses that do not necessarily need to be directly related to the provided cues. Specifically, *capability-exploiting* responses can be induced even by *appearance* cues alone. If the agent looks "too human", users might overestimate its capabilities, and suffer from bad usability. For usable agent-based interaction, each cue must thus be delicately tuned to instill the right beliefs.

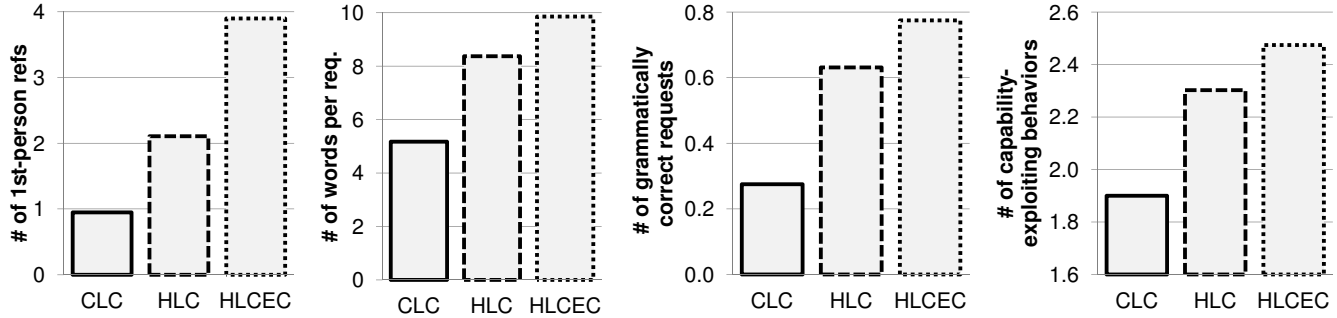

Figure 1. Pps use more human-like and capability-exploiting responses towards systems with human-like cues (HLC) and human-like and capability-exploiting cues (HLCEC), than towards systems with computer-like cues (CLC).

Such fine-tuning projects call for artificial intelligence specialists that can develop smarter systems, social psychologists that know self-presentation techniques, designers that can build these techniques into their characters, and usability researchers that can test the correctness of the formed use image with users. Arguably, only such a multidisciplinary team can bring about a successful paradigm shift from GUIs to agent-based interfaces.

REFERENCES

1. Bradshaw, J. (1997). An Introduction to Software Agents. In J. Bradshaw (Ed.), *Software Agents*, AAAI Press, London, UK.
2. Brennan, S. (1991). Conversation With and Through Computers. *UMUAI 1*, 67-86.
3. Brinkman, W.-P. (2003). *Is usability compositional?* Eindhoven: Technische Universiteit Eindhoven.
4. Chin, J., Diehl, V., and Norman, K. (1988). Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. *CHI'88 Proceedings*, Washington, DC.
5. Chin, M. G., Sims, V. K., Upham Ellis, L., Yordon, R. E., Clark, B. R., Ballion, T., et al. (2005). Developing and Anthropomorphic Tendencies Scale. *HFES Annual Meeting*, Orlando, FL.
6. Cook, J., and Salvendy, G. (1989). Perception of computer dialogue personality: An exploratory study. *IJMMS, 31*, 717- 728.
7. De Laere, K., Lundgren, D., and Howe, S. (1998). The Electronic Mirror: Human-Computer Interaction and Change in Self-Appraisals. *Computers in Human Behavior 14(1)*, 43-59.
8. Dehn, D. M., and van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *IJHCS*, 1-22.
9. Dennett, D. (1987). *The Intentional Stance.* MIT Press.
10. Dey, A. K. (2001). Understanding and Using Context. *Personal and Ubiquitous Computing*, 4-7.
11. Forlizzi, J., Zimmerman, J., Mancuso, V., and Kwak, S. (2007). How Interface Agents Affect Interaction Between Humans and Computers. *Proceedings DPPI 07*, Helsinki, Finland.
12. Halliday, M. A., and Hasan, R. (1976). *Cohesion in English.* Longman Group Ltd., London, UK.
13. Hess, T. J., Fuller, M. A., and Mathew, J. (2005). Involvement and Decision-Making Performance with a Decision Aid: The Influence of Social Multimedia, Gender, and Playfulness. *JMIS 22(3)*, 15-54.
14. Kang, Y-L., Nah, F., and Tan, A-H. (2012). Investigating Intelligent Agents in a 3D Virtual World. *ICIS 2012 proceedings*, Orlando, FL.
15. Laurel, B. (1990). Interface agents: Metaphors With Character. In B. Laurel (Ed.), *The Art of Human-Computer Interface Design.* Addison-Wesley, Reading, MA.
16. Levinson, S. C. (1983). *Pragmatics.* Cambridge University Press, Cambridge, UK.
17. Murray, D., and Bevan, N. (1985). The social psychology of computer conversations. *Proceedings of INTERACT'84,* London, UK.
18. Nickerson, R. (1976). On conversational interaction with computers. *Proceedings of UODIGS '76*, Pittsburgh, PA.
19. Norman, D. A. (1986). Cognivite engineering. In D. A. Norman, and S. Draper (Eds.), *User Centered System Design: New Perspectives on Human-Computer Interaction* (pp. 31-61). Lawrence Erlbaum Ass., Hillsdale, NJ.
20. Nowak, K. L. (2006). The Influence of Anthropomorphism and Agency on Social Judgment in Virtual Environments. *JCMC 9(2)*.
21. Qiu, L., and Benbasat, I. (2009). Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. *JMIS 25(4)*, pp 145-182.
22. Quintanar, L., Crowell, C., and Moskal, P. (1987). The interactive computer as a social stimulus in human-computer interactions. In G. Salvendy, S. Sauter, and J. L. Hurrell, *Social, ergonomic, and stress aspects of work with computers.* Elsevier Science Publishers, Amsterdam, The Netherlands.
23. Reeves, B., and Nass, C. (1996). The Media Equation; How People Treat Computer, Television, and New Media Like Real People and Places. Cambridge University Press, Cambridge, MA.
24. Rosé, C., and Torrey, C. (2005). Interactivity and Expectation: Eliciting Learning Oriented Behavior with Tutorial Dialogue Systems. *INTERACT 2005*, Rome, Italy.
25. Scheutz, M., Cantrell, R., and Schermerhorn, P. (2011). Toward Human-Like Task-based Dialogue Processing for HRI. *AI Magazine 32(4)*, 77-84.
26. Shechtman, N., and Horowitz, L. (2003). Media Inequality in Conversation: How People Behave Differently When Interacting with Computers and People. *Proceedings of CHI 2003*, Ft Lauderdale, FL.
27. Taylor, M. M. (1988). Layered protocol for computer-human dialogue. *IJMMS, 28*, 175-257.
28. Thompson, B. (1980). Linguistic analysis of natural language communication with computers. *Proceedings of COLING 80.* Tokyo.
29. Walker, J., Sproull, L., and Subramani, R. (1994). Using a Human Face in an Interface. *Proceedings of CHI '94*. Boston, MA.
30. Winograd, T. (1972). *Understanding natural language*. Academic Press, Inc., Orlando, FL.