MAKING AGENTS LESS ANNOYING: TOWARDS AN ANIMATED AGENT THAT RESPONDS TO SOCIAL CUES

Marie Gustafsson

Lund University Cognitive Science
Kungshuset, Lundagård
S–222 22 Lund
Sweden

mat99mgu@student.lu.se

ABSTRACT

Animated interface agents have been proposed as a more natural way for the user to interact with computers. Such agents are frequently used in pedagogical and program help applications. While many aspects of social agents have been investigated, that of keeping an appropriate social distance seems to have been overlooked. This ability is coupled with an understanding of when advice is and is not wanted. An agent that gives advice in an Othello program is implemented, which using reinforcement learning adapts its frequency of suggestions to whether or not the user takes the advice. The amount of suggestions that the user can expect is visualized for the user by agent size and position. This agent is compared to one which always gives advice. In the evaluation, test participants experienced that the first agent adapted to their behavior and also found it less distracting.

1 INTRODUCTION

Agents that appear in an embodied, social manner have been a frequent topic of research over the past decade. They are supposed to provide a new way of interacting with computers, and be applicable to education, entertainment, and customer assistance, among others.

Our tendency as humans to anthropomorphize is given as one of the reasons for the need for interaction through social agents, as people already view computers as social beings (Reeves and Nass, 1996). Because of this disposition, it can be assumed that the agent is expected to follow certain social rules. Empirical evidence of the effects of animated agents on user experience and behavior, has been inconclusive and sometimes contradictory (e.g. Dehn and van Mulken, 2000). There has been no consensus on what is important when it comes to agent appearance, personality, feedback, and emotions. The existence of a persona effect has been hypothesized, whereby animated pedagogical agents should make the learning experience more entertaining, making the student more motivated (Lester et al., 1997).

In viewing the interface agents as a social being, attempts have been made to form agent models of personality and emotions to make the agent more believable to the user. One aspect which seems to be overlooked, is that of keeping an appropriate social distance, coupled with understanding when help is wanted and when it is not (which might be something entirely different from when help is and is not needed). If we tell somebody that we don't need their help, they don't turn a somersault for half a minute before disappearing (though perhaps the Microsoft Office Assistant has had to endure enough battering already). It is ironic that in the Lumiére project (Horvitz et al., 1998) from which the Office Assistant is derived, there was a "volume control" in the assistance window, which allowed the user to modify the probability threshold (it uses a Bayesian user model) that determines when the system will provide assistance.

1.1 ON THE JUNGLE OF AGENTS OUT THERE

In the literature on agents, the following terms often appear: social agents, socially intelligent agents, embodied agents, embodied conversational agents, interface agents, animated agents and pedagogical agents. There are probably a few more that can be added to the list.

These appear do be different ways of denoting anthropomorphic agents. Within the broader subject of agents, this aspect is not necessarily present. Further, the term embodied agents can refer to more than one thing; in terms of being an interface and as an approach to artificial intelligence in general. The agent implemented in this paper will be referred to as an animated interface agent, or just animated agent or interface agent.

1.2 Aim of this investigation

I intend to investigate the possibilities for an animated agent to have a more fluid interaction with the user, where the user has direct and indirect control over how much assistance the agent provides. If the user doesn't ask for or doesn't use the aid provided by the agent, then it should step back and disturb the user less. If, on the hand, the user actually uses the suggestion, the agent should become more present or continue to be present. Since I have not found any formal description of this kind of social behavior, it will be categorized as common sense. There are two aims of the investigation. The first is to employ the user's response to the agent's advice as an indicator of how much advice the user wants. The second is to portray the amount of agent advice the user can expect to receive in a tangible manner.

To study these issues, the task to be solved by the user in the investigation should not be complicated, but still require some mental effort. Playing the game of Othello against an invisible computer is satisfactory in meeting these criteria. The agent will give suggestions to the user about what move to make. In one version the agent behaves as described above, while in the other the agent makes a suggestion every time it is the user's turn.

The question of whether or not an embodied agent is desirable for an application is not addressed here. Rather, if the decision is made to include an embodied agent as part of the interface, is it then a good idea to have it adapt to how much the user uses the provided suggestions?

2 Background

The field of interface agents is broad and quite sprawling, with much activity in different directions but little consensus on even the most basic features. The background given here will start out with the field of agents in general, narrowing down to embodied interface agents and why we want to have them. The empirical problems faced by the field is considered, as well as attempts to rigorize it with checklists and frameworks for evaluation. The social perspective on agents requires theories for social intelligence, as well as the breathing of life-likeness into the agent. Finally, the adaptive part of the agent is considered, treating how and to what the agent should adapt, user modeling and reinforcement learning, and the problems with loss of control which adaptivity can cause.

2.1 What is an agent?

As mentioned above, the field of agents is much broader than that of just being used as an interface. It is useful, however, to look at attempts to define this larger field. Dautenhahn (1998) outlines three approaches to defining agents.

- Dictionary definitions that imply that agents have a purpose, exist with other agents, and that there is some relationship between at least two of the agents.
- The view that agency is transient; it is not internal, but attributed. "An object is interpreted as an agent, it cannot 'be' an agent." (Dautenhahn 1998)
- A computational definition, as one discussed by Frankling and Graesser (1996): "An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future."

Different attributes that have been proposed for agents include reactivity, autonomy, collaborative behavior, communicative ability, inferential capability, temporal continuity, personality, adaptivity, and mobility (e.g. Etzioni and Weld, 1995; Franklin and Graesser, 1996). According to Dautenhahn (1998), a well-balanced design of socially intelligent agents needs characteristics of human social intelligence such as embodiment, believability, empathy, the narrative construction of social reality, autobiography and social grounding. Another aspect is that the system should have knowledge both about something and about the user's relation to that something (Negroponte, 1997).

2.2 Embodied interface agents

An agent, as described above, can be used without displaying a graphical, embodied version to the user. An animated interface agent, on the other hand, exhibits various kinds of life-like behaviors, examples of which are emotions, speech, gestures and eye, head and body movements. The focus in interface agents is the user, not the underlying mechanisms of the agent. The important thing is the user's interpretation of the agent as a social being.

Animated agents have been used in a number of systems such as: the PPP-Persona (Rist, André and Müller 1997) for presenting information gathered on the Internet; Herman the Bug (Lester et al., 2001), a pedagogical assistants guiding students through educational programs; and REA, an embodied conversational real-estate agent (Bickmore and Casell, 2000). Perhaps most well-known are the desktop assistants in MS Word, in various incarnations such as a paperclip, Einstein and a kitten.

2.2.1 Why embodied interface agents?

Arguments for the use of embodied interface agents usually assume that this would make the computer system more human-like, engaging, and motivating. With a more human-like system, the interaction might be smoother if standard interaction skills (for example, interpreting facial expressions) were used. There is also the prospect of making the system more entertaining, which might have positive effects on motivation. In the context of education, Elliott, Rickel and Lester (1997) propose that:

> A pedagogical agent with a rich and interesting personality may simply make learning more fun. A student that enjoys interacting with a pedagogical agent will have a more positive perception of the whole learning experience. A student that enjoys a learning environment will undoubtedly spend more time there, which is likely to increase learning.

Arguments against the use of embodied interface agents are that it may lead to false mental models of the system (Norman, 1997), such as assumptions that the agent's emotional and cognitive funcions are more like those as humans than they really are, thus misleading expectations. Another concern with embodied agents is that they may attract too much attention, becoming a distraction (Walker, Sproull, and Subramani, 1994). Further, interface agents may increase user anxiety, reduce user control, undermine user responsibility, and weaken the user's sense of accomplishment (Shneiderman and Maes, 1997). Maes (1997) explains that the agents in the presented mail filtering program were deliberately drawn as simple cartoon faces, to avoid attribution of human-like intelligence.

There are not yet any conclusive empirical evidence to support either side. Dehn and van Mulken (2000) outline different experimental approaches used in research on the effects of animated agents in human-computer interaction:

- the kind of animation used, which relates to the realization of the experimental treatment
- the kind of comparison made, relating to the choice of control condition
- which measures are taken as indicators of an effect, which concerns the choice of dependent variable
- the task to be performed while interacting with the system, relating to the choice of experimental setting.

Because of these differences it is difficult to get an overview of the empirical evidence and even more difficult to integrate as a whole.

2.2.2 Difficulties in developing and evaluating animated agents

There is an abundance of different research issues involved in the field of interface agents. Many different prototypes have been built, in different domains and according to different paradigms, making it difficult to assess and compare embodied agents. Research papers on embodied agents often describe the construction of a complete prototype. Discussions about new application areas and interaction techniques are mixed with those of technical innovations. Appearances, personality, and behaviors are often a result of introspection and personal preferences. Isbister and Doyle point out that this is "not a result of flawed research" (2002), but because the field is new and in an explorative stage. They do point out that there is a need to develop criteria for design and evaluation and also propose a taxonomy of research areas within the field of embodied conversational agents. The different categories they propose are:

- believability, where example properties are emotion, strong personalities, variability in movement and response, personalizability, appearing to care what happens
- social interface, where example properties are social context, social behavior, knowledge of other agents, ability to cooperate
- application domains, where example properties are domain knowledge, contextuality, timeliness, risk and trust
- agency and computational issues, where example properties are autonomy, responsiveness, reactivity, reliability, completeness and efficiency
- production, where examples of properties are professional and consistent quality and integration of the whole

For each of these categories Isbister and Doyle propose criteria for success and evaluation techniques. The topic of interest for this paper would mostly fall under social interface, where the proposed evaluation techniques are "qualitative measures from user of agent's friendliness, helpfulness, or intuitive communication ability; quantitive measure of speed, ease, satisfaction with achievement of task"

Xiao, Stasko and Catrambone (2002) also conclude that there has not been much careful empirical evaluation of embodied conversational agents (ECA). According to them, whether or not ECA interfaces are useful or not is too general a question, as it depends on the agent's specific behaviors, the user's characteristics, and the nature of the task the user has to perform. A framework for evaluation ECA is proposed, along the dimensions of features of the user (for example, personality, background knowledge, and goal), features of the agent (for example visual appearance, personality, presence and initiative), and features of the task (for example, domain, objectiveness, and intent). Dimensions of usefulness are performance and satisfaction.

Ruttkay, Dormann and Noot (2002) provide a checklist to compare embodied conversational agents, from the points of view of design, usability, practical usage and user perception. The design aspect includes the embodiement of the agent, the mind of the agent, control and interactivity, as well as the application context. In usability, proposed evaluation dimensions are learnability/ease of use, efficiency, and errors. Evaluation of user perception includes helpfulness, user's satisfaction, believability, trust, and engagement. The conditions and resources are important to the practical applicability dimension. For usability and user perception aspects, the authors provide a definition and discussion of each subaspect.

Dehn and van Mulken (2000) categorize the proposed possible effects of animated agents as

- improvements in the user's subjective experience, where the system might be perceived as more entertaining, informative or interesting than one without an animated agent.

- changes in user behaviour while interacting with the system in a beneficial way

- performance data may indicate an improved outcome of the interaction.

One problem of much of the research on interaction with embodied, social agent systems has been, according to Cassell and Thórisson (1999), the lack of computers which can support multimodal dialogue with a human user. These older studies have used various methods to avoid this problem and still investigate lifelikeness, trust, effectiveness of communication and user's likability of the interaction. According to Cassel and Thórisson, "one cannot justifiably generalize the results of these studies and systems to future systems employing computer-controlled characters capable of realtime dialogue."

Höök, Persson and Sjölinder (2000) aimed to develop "new methods and criteria for measuring users' understanding of, and socio-emotional reactions to, believable characters." The methods they used in testing their Agneta & Frida program were videotaping the users' face, measuring the time spent with the user's estimation of time spent, together with questions about anthropomorphism and believability.

Much research in the domain of interface agents is concentrated on user reactions, as other aspects, such as effects on learning, are more difficult to measure. However, self reports are not necessarily a true representation of user reactions. In attempting to evaluate embodied conversational agents, one could compare their abilities to those of humans. But since we do not understand the intricacies of communication between humans, this is difficult. Another is to rely on more traditional usability tests. But what should be measured, and how? The things that come to mind to measure, such as believability, ease of use, and perceived intelligence, are concepts that are in themselves rather hard to define.

2.3 Social interaction with agents

Building on studies such as the ones presented by Reeves and Nass (1996), there is an assumption that since people naturally apply social rules to interaction with computers, this should be used in interaction design.

2.3.1 Anthropomorphism

Reeves and Nass (1996) argue that since humans have both evolved and learned a certain system for interaction with other intelligences a system of social interaction we naturally attempt to employ that system in dealing with computers. They provide several experiments where people have treated computers as people. Laurel (1997) points out that the tasks computers perform for us require that they express two distinctly anthropomorphic qualities: responsiveness and capacity to perform actions.

2.3.2 Theorizing about social intelligence

According to Gong (2002), some fundamental mechanisms in social interaction are overlooked in interface agents. In his attempt to theorize about social intelligence, four principles are identified: attractive, affective, appropriate, and adaptive. Gong is focused on the attractive and affective aspects, especially on that in order to get a positive user response, the face of the agent should be attractive. Appropriateness acts as an overarching general rule. Behaviors such as verbal and emotional expressions should be appropriate to its role, target users and contexts. Adaptability is considered an essential factor in communication competence and social functioning (Duran, 1992). Duran conceptualizes communicative adaptability as "the ability to perceive socio-interpersonal relationships and adapt one's behaviors and goals accordingly."

People have a need for personal space, which can be categorized into four types: intimate distance, personal space, social distance and public distance, which range from very close to further away from the person. It has been found that violation of personal space leads to discomfort (Smither, 1998).

2.3.3 BELIEVABILITY, EMOTION AND PERSONALITY

One of the objectives of designing an animated interface is to get a positive user experience of the system, and it is believed that this can be achieved by making the agent more life-like. The dimensions addressed are often those of intelligence, usefulness and how comfortable the interaction was.

An all encompassing concept for the perception of the agent is that of believability, which was introduced to the domain by Joseph Bates (1994). The idea is to give an ’illusion of life’. Drawing on experiences from literature and animated movies, among others, Bates concludes that that the portrayal of emotion is very important to believability. Emotions show that the agent has desires and is concerned about what happens in the world, and “if they don’t care, neither will we.” However, according to Bates it is not required that the emotional state is represented in any obvious way within the agent, as the objective is merely to create an illusion of the emotion.

In the agent literature as a whole, however, what believability is to mean and what effects it is supposed to have on the user is rather unclear. Some aspects which are believed to be part of the concept are facial expressions and body language, personality and attitude (Höök, 2000 et al., among others). Dehn and van Mulken (2000) assume that users will find the system believable if they perceives it as intelligent and competent.

According to Dautenhahn (1998) the social agent “need not necessarily appear or act just like biological agents, but some aspect in their behavior has to be natural, appealing, life-like” . Further, she concludes that “believability is in the eye of the observer which means that it is influenced by the observer’s individual personality, naive psychology and empathy mechanisms” (Dautenhahn, 1997). As Porter and Susman (2000) put it: “Life-like does not mean ‘has movement’; life-like means ‘has a brain’.”

These concepts are also somewhat tied into the appearance of the agent; how realistic or human-like should it be? There has been no conclusive empirical evidence on the subject. Further, the domain needs to be considered, as a cartoon-style agent may be appropriate in a game but not in a business application.

2.3.4 WHEN DO WE WANT A SOCIAL AGENT?

Doyle (1999) outlines several situations when a communicative agent might be a good idea. One of the aspects is whether the user is uncertain or decisive. If the user needs to make choices, especially choices about which the user is unsure, then a communicative agent can be useful. If the user knows how to accomplish a task then direct-manipulation is faster and more transparent. Social interaction is anything but transparent, requiring explanations, clarifications, and redundancy, which are valuable in teaching, but not if you know how to perform a task.

The domain of the application is an important factor when considering the effect of an interface agent, as pointed out by Dehn and van Mulken (2000). Further, they propose that “an animated agent can only be expected to improve human-computer interaction, if it shows some behavior that is functional with regard to the system’s aim” When the agent’s behavior “neither maps onto system behavior nor conveys any other information that is relevant”, such as behavior displayed during idle time to make the agent appear more lively, might attract the user’s attention even though there is no information provided. A fitting analogy for badly designed social agents is that of backseat drivers, constantly making annoying suggestions (Etzioni and Weld, 1995).

2.4 AGENT ADAPTIVITY

Adaptivity is one of the basic attributes proposed for agents. What should an adaptive agent do? According to Erickson (1997), systems with adaptive functionality are in general doing three things:

- Noticing: trying to detect potentially relevant cues
- Interpreting: trying to recognize the events by applying a set of recognition rules
- Responding: acting on the interpreted events by using a set of action rules either by taking some action that affects the user, or by altering their own rules (i.e. learning)

These are similar to the issues that Middleton (2001) believes must be addressed for collaboration with the user to be successful: knowing the user; interacting with the user; and competence in helping the user. However, none of these are easily achieved. To know the user, the user’s goals and intentions must be extracted, these must be put into a context and they might change. For example, the user will typically be working on several concurrent tasks. Further, the initial training time must be reduced, as users will want results early. Maes (1997) lists four sources for how the agent might learn from the user’s actions:

- by “looking over the shoulder” of the user
- direct and indirect (ignoring agent’s advice) user feedback
- from examples provided explicitly by user
- ask advice from other agents with same task for other users

Among the challenges of interaction are amount of control to delegate and building trust (Middleton, 2001). As for the issue of competency, the agent needs to know when (and if) to interrupt the user, perform tasks autonomously in the user's preferred way, and finding strategies for partial automation of tasks.

2.4.1 Methods for making the agent adaptive

One way of making the agent adapt to how the user interacts with the system is for the agent to have a model of the user. The field of user modeling has been around for about twenty years, starting with student modeling in the early eighties. After this early domination for student modeling, interest was turned to the demands of electronic commerce and the field of information retrieval. The cognitive processes that underlie the user's actions and the user's behavioral patterns or preferences are some things that user models may wish to describe. Another is the difference between the user's skills and expert skills (Webb, Pazzani and Catrambone, 2001).

Among others, the following structures and processes are often included in a user modeling system (Kobsa, 2001):

- the representation of assumptions about user characteristics in models of individual users, such as assumptions about knowledge, misconceptions, goals and preferences;
- the representation of common characteristics of users, grouping them into subgroups, or stereotypes;
- the classification of users as belonging to one or more subgroups, along with the integration of typical characteristics of these subgroups into the current individual user model;
- the recording of users' behavior, especially their past interaction with the system;
- the formation of assumptions about the user based on the interaction history.

Standard machine learning techniques seem to be a good candidate for user modeling. Observing the user's behavior can provide training examples, which a machine learning method can use to make a model to predict future actions. But there are several problems in applying machine learning to user modeling: the need for large data sets; the need for labeled data; concept drift; and computational complexity (Webb et al., 2001). Further, if user modeling profiles are to be created, a sufficient amount of time is required before they can be of any use. A solution to this might be the incorporation of stereotypes. Ideally, these should be used for initializing the user model, until there is more information about the individual user. Also, it should be regularly checked whether or not the right stereotype is activated (Virvou and Kabassi, 2002).

The user model easily becomes quite complex and needs to represent many forms of data about the user. A seemingly simpler way of achieving a responsiveness, albeit without as much detail, is through reinforcement learning, an artificial intelligence technique adapted from ideas in animal-learning theory. In reinforcement learning, the agent must learn behavior through trial-and-error interactions with a dynamic environment. In the standard model, an agent is connected to the environment via perception and action. For each interaction step, the agent's input is some indication of the current state of the environment, whereby the agent chooses some action as output. The action changes the state of the environment, and the value of this transition is communicated to the agent as a reinforcement signal (Kaebling and Moore, 1996).

2.4.2 Problems with adaptive agent behavior

We want agents to act in an intelligent manner, but Baylor (2000) points out that although the agent should seem competent to the learner, it should not be too intelligent if it is to be effective. If it is, it may lead to "unrealistic expectations, a loss of control and limited understanding as to the agent's reasoning" (Baylor, 2000).

Further, when the agent is trying to confirm whether or not the advice it gave was useful, it will probably check to see if the user followed it. Horvitz et al. (1998) describes a Wizard of Oz setup where Excel experts were to figure out what task a user was trying to complete and help with it, by only looking at the screen as this was being done. To help out, they could offer advice, that the user thought was from the computer. Poor advice could be quite costly to users. Often, participants would become distracted by the advice and start to experiment with the features described by the wizard, which in turn would give experts false confirmation of successful goal recognition.

Another problem with agent adaptivity is the possiblity of decreased user control (Schneiderman and Maes, 1997). In taking away seemingly irrelevant information, the user might become confused (Chin, 2001). If we take the case of a pedagogical agent, which is part of the planning of the instructional environment, how should this be communicated to the learner? As Erickson (1997) puts it:

> Consider an intelligent tutoring system that is teaching introductory physics to a teenager. Suppose the system notices that the student learns best when information is presented as diagrams and adapts its presentation appropriately. But even as the system is watching for events, interpreting them, and adjusting its actions, so is the student watching the system, and trying to interpret what the system is doing. Suppose that after a while the stu-

> dent notices that the presentation consists of diagrams rather than equations: it is likely that the student will wonder why: " Does the system think I'm stupid? If I start to do better, will it present me with equations again?" There is no guarantee that the student's interpretations will correspond with the system's.

Erickson therefore believes that if we want to make agents that interact smoothly with the user, we need more research focused on the portrayal of adaptive functionality, rather than the functionality itself.

2.4.3 Evaluating user models and adaptive behavior

How are user models and user-adapted systems to be empirically evaluated? In Chin (2001), which deals with the subject, it is pointed out that it is likely that the participants will be biased by the mere belief that a user model or other advanced technology is present in the program. Questions that should be asked are whether the user model adaptations actually improve the system and what types of users benefit from the adaptations. To some classes of users, it may be that case that some adaptations are less beneficial or even detrimental. Another issue is that when adding a user model, in general the software system will become more complex, less predictable, and more buggy.

2.5 Agent advice, user feedback and control

Feedback and advice from the agent can be both unsolicited and user-initiated. There has been little research on the best amount and timing of the advice. Baylor (2002) proposes that:

> the principle of minimal help could allow the student to select a feedback option depending on the amount of structure, interaction, and feedback s/he desires when problem-solving. If the agent could fade and allow more student initiative as the student gains expertise, it could also address this issue.

Also, it is important that the agent doesn't take over thinking for the student (Baylor, 2000). As an overall guideline she proposes that the agent should not give feedback that is too explicit, and that ideally, some way for the learner to control the amount of feedback should be provided by the system (Baylor, 2001).

Dufresne and Hudon (2002) discusses that there is a need for the user to be able to personalize the interaction with the agent, but that just setting preferences might become tedious. They intend to integrate features into the learning environment that take into account the user's reactions to support. Among the personalizations are the possiblities to choose from various coaches, with different personalities, to ask for more or less support and also to chose whether the agent should use humor or not.

In the Agneta & Frida program made by Höök et al. (2000), the user could set the level of activity of the characters from 0 to 5. Unfortunately, they do not report on how this function was used by the test participants. Laurel (1997) proposes that the traits of agents be fully user-configurable, and that the option to not have an agent should be possible.

Arguing for more user-control, Malone, Lai and Grant (1997) propose a semi-formal system. "Don't build agents that try to figure out for themselves things that humans could easily tell them." This is to avoid, among other things, that the agent will infer the incorrect rule when the user might easily have provided the system with the rule they wanted.

The field of interface agents is wide, with many important parts in which no consensus has been reached about what makes an effective agent, nor is there any criteria for what impact an effective agent has on the user. However, the amount of research put into the field shows that many see a potential in it, though the potential they see may be different. Some see it as a possibility to improve human-computer interaction, some as a way to improve pedagogical programs, others as a method to better understand human cognition, and yet others as means for making lucrative applications. Key parts are how to portray the agent to the user, how to make the interaction smooth, and how to decide what, when and how for agent advice.

3 The Program

To test user reactions to a more socially responsive pedagogical agent, two versions of an Othello program were made. In the control, the static version, the agent always gives advice, with no consideration to whether the user follows it or not. In the adaptive version, the user gives indirect feedback by either using or not using the agent's advice, and gives direct feedback by asking the agent to be more or less present.

3.1 The Othello intelligence

For a background on programming a game intelligence, see Russel and Norvig (1995), where games as searches, evaluation functions, and alpha-beta searches are covered. The Othello part of the program uses an alpha-beta search with a positional evaluation. The search has depth five, which might be a bit shallow, but with deeper searches it took noticeably (half a second to a second) longer. It might have been better if an iterative deepening search had been used. The positional evaluation worked by adding up the values of the positions held by the player. Highest value was given for the corners, while negative values were given for the squares

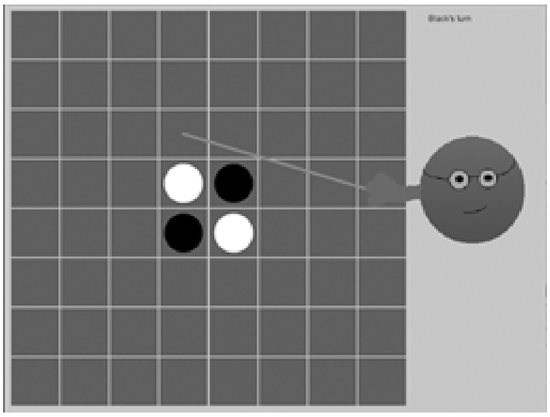


FIGURE 1: *The game board and the agent, as seen by the user when the program is first started.*

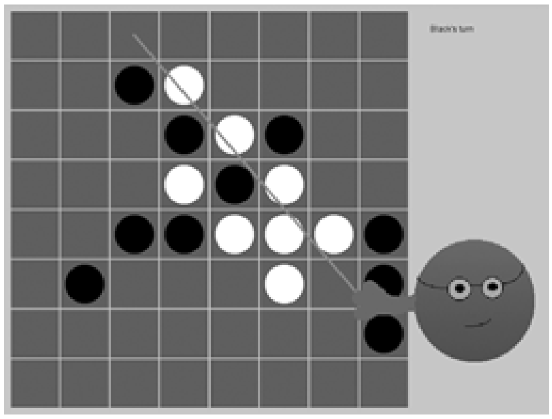


FIGURE 2: *Here the agent is very far forward, and gives advice almost every turn. I.e. the iquote is high.*

adjacent to the corners. The other squares on the edges had relatively high values, and all other positions had a small positive value. Both the agent and the computer opponent used the same search depth and evaluation.

3.2 WHAT SHOULD THE AGENT LOOK LIKE?

Agent appearance, such as the spectra between realistic and iconic and expression of emotional and envelope feedback is a big question in itself. Since this experiment wishes to test the effect on user of agent responsiveness to user feedback, the ideal would be to keep aspects such as agent appearance neutral. But there is currently no consensus on what neutral is in this case (or if it is even possible or desired). Therefore the agent's appearance will be kept very simple: it is a blue circle with some basic facial features (mouth leaning towards a smile and eyes behind a pair of pink glasses). It has only one hand, which facing the game board and holds a pointer (see Figure 1).

3.3 WHAT THE USER ENCOUNTERS

When the program is started, the user sees a window with an Othello game board in starting position. The agent is to the right of the board. It offers advice on the current move by pointing to a square on the board. For the first move, the agent always gives advice.

In one version of the program, the agent is static and gives advice every time it is the user's turn. In the responsive version, whether or not the user takes the advice affects the agent. If the user takes the advice, the agent gets positive indirect feedback and moves forward. That is, it gets slightly larger and moves a little toward the bottom of the window. If the user doesn't take the advice, the agent gets negative indirect feedback and moves backward, i.e. gets smaller and moves up.

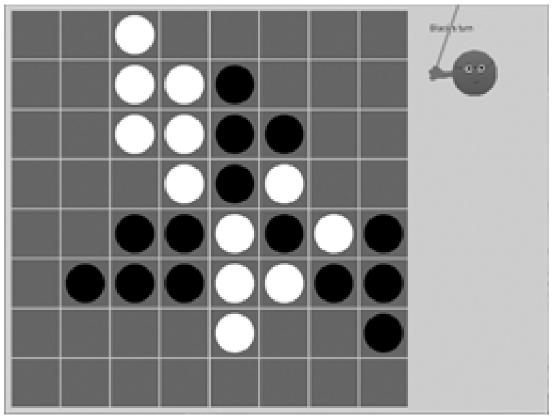


FIGURE 3: *Here the agent is in the opposite position to Figure 2: it is very far back and almost never gives advice. Hence it has a low iquote.*

Further, in the responsive version, the user can provide direct feedback by dragging the agent down or up, i.e. forward or backward. When the cursor is placed on the agent it turns into a hand (otherwise it is just the usual arrow). If the user clicks the mouse button, the hand turns into a 'grabbing' hand, indicating to the user that s/he has picked up the agent and by holding down the button, s/he can move it to the desired position.

Thus, the agent's vertical position (it cannot be move horizontally) determines how much advice it gives. When it is at the bottom it always gives advice (see Figure 2), while it never gives advice when it is at the top (see Figure 3). In between these two positions, it is more probable that advice will be give the closer the agent gets to the bottom.

This setup is quite similar to the "volume control" proposed by Horvitz et al. (1998) which, as mentioned in the introduction, allowed the user to modify the probability threshold determining when and if the system will provide assistance. Here the agent can be said to embody this volume control (though the underlying system for determining advice is different).

3.4 The adaptive behavior

Reinforcement learning is used to control how much advice the agent gives. Positive feedback is given if the user follows the advice, negative if s/he does not. More specifically, what determines the agents frequency of suggested moves is its interaction quote, or *iquote*, which is a number between zero and one hundred. When it is its smallest the *iquote* is zero, when largest one hundred. When it is the user's turn, whether or not to provide a suggestion is determined randomly using the *iquote*. If it is zero, no suggestion is given, when it is one hundred, a suggestion is always given, and in between those it becomes more probable the higher the *iquote* is. Every time the agent provides a suggestion, the *iquote* is updated based on whether the advice is followed or not. If it is followed, the agent receives positive feedback:

$$iquote = iquote + (100 - iquote) * .3$$

And if it is not, the agent gets negative feedback:

$$iquote = iquote - (iquote) * .3$$

This rather simple way of updating the iquote provides rather large changes when positive feedback is given to an agent with a low iquote and when negative feedback is given to an agent with a high iquote. Positive feedback when the iquote is high or negative feedback when the iquote is low has less dramatic effects. This way of updating the iquote was chosen to give quick responses to the user's actions, while trying to prevent that the agent comes too far towards either extreme.

The only user model the agent has is the iquote. This is an aggregation of user feedback, and it adapts quickly to changes in user behavior. This straightforward method is possible as no attempt is made to understand how the user view's the game, only how the user views the advice from the agent.

The adaptive behavior described above deals with indirect feedback. As described, the user can also give feedback more directly to the agent, by dragging the agent up or down. When this is done, the iquote is updated correspondingly; the iquote and the agent's vertical position vary linearly.

3.5 Implementation details

Everything except the Othello intelligence was implemented in Macromedia Director. More specifically, this included the graphic representation, the game board and it's updating, as well as the agent's behavior. The reason for using Director is the ease with which the interface could be made and manipulated. For the Othello intelligence (which was originally constructed for another purpose than this experiment) java was used. The director and java programs communicated via a TCP socket.

4 Experimental design

4.1 Hypotheses

It is expected that the user will feel more comfortable with the version of the agent that adapts to how many of the suggestions that the user follows. More specifically, it is expected that

(H1) the participant will feel that the adaptive version adjusted more to their behavior than the static version did

(H2) the participant will feel that the interaction with the adaptive version was better than the interaction with the static version

(H3) the participant will find the adaptive version to be less distracting than the static version

(H4) the overall impression of the adaptive version will be better than that of the static version

4.2 Participants

After a pilot test on one person, the results of whom were not included in the results, a total of sixteen people tested and evaluated the program, nine women and seven men. They were aged between 21 and 31, with a mean of 24. All were currently enrolled in or had completed a university education, and computer experience varied from frequent use of basic applications such as word processors, email and Internet programs, to more specialized ones and to web design and programming. Seven of them had or were in the midst of getting a computer related education, eight said they used basic applications often, and one person used them sometimes. Most of them were familiar with the rules of Othello, a few had all but forgotten them while a few others were very experienced with Othello and other strategic games.

4.3 Tasks and Procedures

Each participant tested both the stationary and non-feedback agent and the responsive, feedback taking agent, though the order in which they were tested varied. The tests were carried out in a calm setting such as an office or a kitchen, in the participant's or author's home. If the participant felt uncertain about the rules of the game, a paper with basic Othello rules was provided. The person being tested was provided with instructions (see Appendix A), played a round of Othello and filled out a questionnaire (see Appendix B), which was then repeated for the other version of the game. After the second game they also wrote answers to some more open ended questions about previous experience with agents, what was better and worse with this agent, as well as about the appearance and mimicry of the

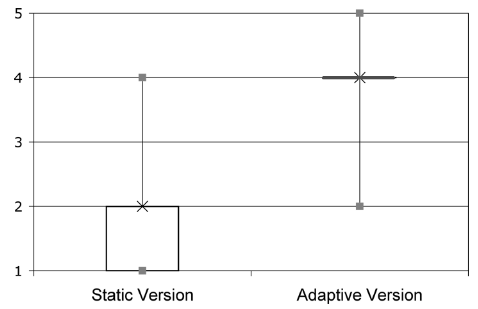


FIGURE 5: *Boxplot of the test participants' responses to the question "Did you experience that the agent adapted to your behavior?" The X indicates the median.*

agent (for the whole list, see Appendix B). A brief informal interview with debriefing was carried out when they were finished. Apart from user responses to questionnaires, the program recorded the time the user spent with the program, who won, how many suggestions the agent offered as well as the number of suggestions that the user followed.

5 RESULTS

The questionnaire that the test participants responded to contained eight questions, which were answered on a five point Likert scale (where 1 = none/bad and 5 = a lot/good). Four of these pertained directly to the hypotheses, and will be treated separately below. Differences based on whether the participant has a computer related education or not, and bases on gender will also be noted. As for experience with Othello, since most people were familiar with the rules of the game but not much more, differences based on this will not be noted.

5.1 DIFFERENCES IN EXPERIENCED ADAPTABILITY (H1)

This hypothesis pertains to the question "Did you experience that the agent adapted to your behavior?" (question four). For the static version, the mean was 1.81 (sd = 0.91). For the adaptive version, the mean was 3.94 (sd = 0.77). A paired t-test gives t = 7.27, which is significant at $p < 0.01$ (one-tailed). Thus hypothesis one can be confirmed, there was a significant difference between the means. See Figure 4(a).

There were no significant differences based on computer experience or gender.

5.2 DIFFERENCES IN INTERACTION (H2)

The hypothesis relates to the question "How good would you say that the interaction between you and the agent was?" (question one). For the static version, the mean was 2.88 (sd = 0.72). For the adaptive version, the mean was 3.38 (sd = 0.72). A paired t-test gives t = 1.65, which is not significant (two-tailed). There is not enough support to confirm hypothesis two. See Figure 4(b).

There was no significant difference between those who had a computer related education and those who did not, though those with a computer related education (seven of the participants) experienced a larger positive difference between the adaptive version and the static version (mean of +1) than those without (mean of 0.11). Further, a few of those with no computer related education rated the adaptive version lower than the static version (three out of nine of those without computer education rated the adaptive version one point lower, and two rated them the same; for those with computer education, one rated the adaptive version one point lower, and two rated them the same). This could be because of the word interaction, which is not well established in Swedish outside human-computer interaction, which became apparent during testing as many test participants were uncertain of what the question alluded to.

Fewer women than men saw any positive difference in interaction between static version and adaptive version. This could be because only two out of nine women had a computer related education, while five out of seven men did.

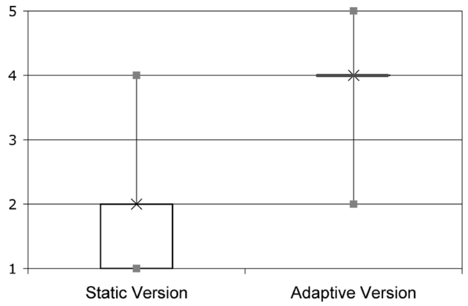
TABLE 1: *Means and standard deviations for user responses to the question "Did you find the agent distracting?", separating women's and men's responses.*

		Static Version	Adaptive Version
Women	mean	3.0	2.11
	s.d.	1.11	0.93
Men	mean	1.57	1.29
	s.d.	1.13	0.49

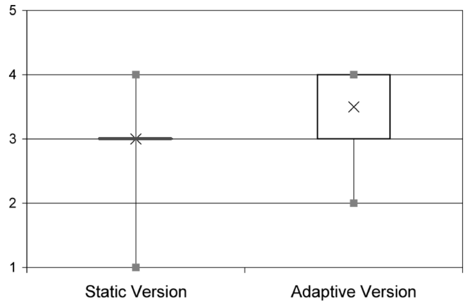
5.3 HOW DISTRACTING WERE THE AGENTS? (H3)

The participants' question for this hypothesis was "Did you find the agent distracting?" (question five). For the static version, the mean was 2.38 (sd = 1.31). For the adaptive version, the mean was 1.75 (sd = 0.86). A paired t-test gives t = 2.95, which is significant at $p < 0.01$ (two-tailed), which confirms hypothesis three. See Figure 4(c). There was no significant difference when it came to computer education.

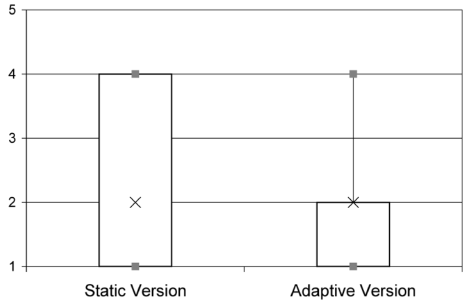
For gender, there was a significant difference both for the scores for the static version (p = 0.03) and for the scores for the adaptive version (p = 0.04). There was also a significant difference for the difference between the static version and the adaptive version for women, but not for men. This is because most of the men did not find either version to be distracting, while women found the adaptive version to be less distracting than the static version (see Table 1.



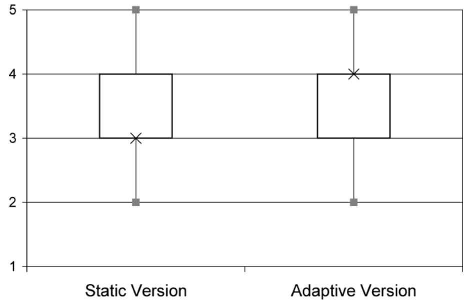
(a) “Did you experience that the agent adapted to your behavior?” (H1)



(b) “How good would you say that the interaction between you and the agent was?” (H2)



(c) “Did you find the agent distracting?”(H3)



(d) “What was your overall impression of the agent?”(H4)

FIGURE 4: *Boxplot for the participants responses to questions corresponding to hypotheses one through four. The X indicates the median.*

5.4 THE OVERALL IMPRESSION (H4)

The hypothesis corresponds to the question “What was your overall impression of the agent?” (question eight). For the static version, the mean was 3.31 (sd = 0.79). For the adaptive version, the mean was 3.69 (sd = 0.87). A paired t-test gives t = 1.69, which is not significant (two-tailed), thus hypothesis four cannot be confirmed. See Figure 4(d).

There were no effects from either gender or computer related education.

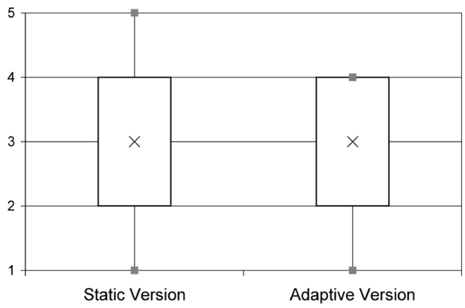


FIGURE 6: *Boxplot of the test participants’ responses to the question “Did you find that the agent behaved in a predictable way?”. The X marks the median.*

5.5 AGENT PREDICTABILITY

One concern with the agent that varied the amount of advice was that it might have been viewed as more unpredictable. The replies to the questionnaires give no indication of this. For the static version, the mean was 3.06 (sd = 1.06). For the adaptive version, the mean was 2.94 (sd = 1.12). So, while the agent cannot be said to be unpredictable, it cannot be said to be predictable either. Further, it is possible that some test participants applied the predictability question to the suggestions themselves, rather than to when they were supposed to come. Perhaps the question should have been worded differently. The word behaved is rather general, and it might be best if the question was split up into several, and have the participant answer with his/her own words on how they perceived the mapping of amount of advice given to position, and whether it was negative not to know whether or not advice was going to be received.

5.6 OTHER POINTS OF INTEREST

Apart from the questionnaire, some data was recorded while the participant played: duration of the game, number of suggestions made by the agent, how many of these that the player followed, and who won. There seems to be no correlation between the time spent with the game and user reactions. However, a greater amount of time was spent with the first version played. Neither do the amount of advice given and taken seem to have affected the ratings.

Several problems were encountered during testing. One was, as already mentioned, that in Swedish the word interaction is not used by people in general. Another word should probably have been chosen. Another was that the test participants did not always read the instructions very well, and thus some did not really notice the agent's change in behavior, even though it was rather explicitly explained in the instructions.

6 Discussion

The purpose of this experiment was to see how users react to an agent which adapts to whether or not the user takes the advice. No attempt is being made to see if this improves learning. Rather, the aim was to find a way to step around the often occuring annoyance effect with pedagogical agents. Further, the graphical representation of the agent as being smaller or larger according to how much advice it will give is a promising way of giving the user more insight into how the agent's adaptivity works.

This metaphor is not without problems, though. It relies on ocular cues that tell us that closer objects appear larger and further down from the horizon than those that are farther away. But there are also metaphors based on "more is up, less is down" (Lakoff and Johnson, 1980). Many preference settings on computers work this way. Also, in Macintosh operating systems, the hard drive is placed in the upper right hand corner. The user might be less inclined to believe that the agent is less important when it is in this position.

The results indicate that the test participants experienced that the agent adapted its amount of advice to them. Further, they found the responsive agent less distracting. The scores for viewing the interaction with and the overall impression of the adaptive version as better were not significant.

In an attempt to keep other factors than the social responsiveness neutral, the agent appearance was kept very simple, and no attempt was made to make the agent appear more life-like by adding personality or emotions. Also, given the time and scope of this investigation, this was not an option. However, these aspects could be very important to the user's perception of the agent, and it is possible that the absence of these characteristics might have affected the outcome more than their precense would have. For example, Gong (2002) proposes that the agent must have an attractive appearance to get a positive user response.

It is quite possible that the lack of these other aspects might have had a slightly detrimental effect on the test participants' overall impression of the agent, which might have affected the answers to the questionnaires. This effect might be applicable for the experience as a whole. These are issues that all research in this area face, as it is difficult if not impossible to keep certain factors neutral. In researching one aspect of interface agents, how many other aspects need to be included to get reliable results?

The participants were asked explicitly about the agent's appearance and answers to this varied greatly. Responses ranged from cute, simple (both as something good and as something bad), childish, okay, to ugly. Maybe discussions and research on what level of iconization should be applied to the agent are asking the wrong questions. Such preferences seem to be rather individual. After all, people's opinions about what is visually pleasing usually are.

One participant asked why the agent had to be male. I had not thought about that, I think I was aiming at something slightly gender neutral. In looking over the responses in which the participants refer the agent, all who do write he. Also, it is blue; had it been pink it might have been considered female. Or maybe it is that the neutral is inferred as male, and to be viewed as female certain female attributes (such as a bow and long eyelashes) have to be added. The computer game Ms. Packman is a good example of this.

The expression on the agent's face in the program might be described as a sly smile. In an older version, used for the pilot participant, the facial expression had the mouth turned more downwards, but it was judged as a bit too negative.

The test participants were asked if and how they would like the agent's facial expression to vary. Out of those who replied to the question, a few were outright negative. Some more were positive, with motivations such as that it would make the interaction stronger, would be fun, and would give it the sense of being a team. Most respondents could not quite make up their minds. One thought that while varied facial expressions might be fun, it might get annoying and distracting. One thought that a few more expressions might be good, but that none should be negative. A few users suggested that the agent's expression could vary with how well the user is doing (in Othello this might be a bit difficult as evaluation of the game board is among the trickiest parts of making an Othello program).

Almost all of the participants had the Microsoft Office Assistant as their only previous experience of interface agents. A few had used agents in computer games, where they had the role of teaching the game basics. The test persons were asked to compare the agent in the Othello game to those they had used previously, giving both what was better and what was worse. Things that were better was that it was not as obtrusive and annoying as the paperclip, that it gave relevant advice, that it is easily inactivated and then brought back again, and that it is not in the way. Things that were worse were that it could not be turned off, that it was a bit impersonal, that it wasn't very good at playing Othello, and that it would have been better if it had provided more information. Adding the option to ask the agent to motivate its suggestion is an interesting possibility. In the case of the Othello game, this could be a succession of a few Othello boards representing how the agent thinks that the game could advance the next five moves in response to the suggested move. It could also explain why certain moves were not recommended.

Only two of the participants managed to win over the computer player. A few complained that the advice given by the agent was a bit stupid, and one person came to the conclusion that the agent gave suggestions that were to the computer player's advantage, thus trying to trick the user. The agent player and the computer player have the same game algorithm, so this effect was a bit surprising. It is possible that the program plays white better, or that it plays better when it does not have the first move. This is a most unfortunate defect, as the credibility of the agent most probably decreases if the user doesn't perceive the agent's advice as good.

Another thing that might have given more useful results would have been to use younger participants for the test, who might be a more appropriate age group when it comes to actually using an agent of this sort. Also, the test persons could have played with the agent for a longer time period, getting more acquainted with the agent's behavior. There is also the problem of a politeness bias, which is probably especially problematic in this case, where the participants knew the author.

One possible extension of this agent is having the different positions of the agent correspond more directly to the amount of agent advice. In the implementation used here, there is some randomness to it. One way of doing it is to have the suggestions prioritized, and the agent position correspond to how high the priority has to be for the advice to be presented. It is easy, but perhaps not correct, to make assumptions about where the user would place the agent depending on his or her skill. One such assumption is that the novice would prefer to get a lot of advice, while the more experienced user would place the agent further back. In such a way, it could be inferred that a more withdrawn agent could be appreciated when offering more complicated or esoteric advice.

The Othello agent used in testing did not adapt the amount of advice to how well the user is doing. Unfortunately, in Othello, the game state is difficult to evaluate board evaluation is one of the most difficult parts in making an Othello program. In a learning situation, it might be desirable that the agent provide less advice as the learner becomes more adept at the task (as suggested by, for example, Baylor, 2002). This could be visualized as the agent taking a step back as the Othello agent, but instead of being responsive to whether the user takes the agent's advice, it could be responsive to the number of errors made by the learner (or a combination of the two). Adjusting advice the learner's needs is not a new idea, but the visualization might be a helpful cue. Keeping down the factors taken into account might make it easier for the user to understand what the agent is doing, which taps into the discussion of "glass boxes" vs. "black boxes" (in e.g. Malone et al., 1997).

In the implemented program, there was no way to lock the agent in a certain position. As soon as the user took or did not take the agent's advice, the agent updated it's position. It would be desirable to have some added functionality so that the user could tell the agent to stand still. Another troublesome matter is that of knowing when the user has taken the agent's advice. It is straightforward in Othello, but there are many domains where this is not the case. Also, the user might take the advice at some later point in time - should this reinforce the agent? Will the user understand why? (e.g. Middelton, 2001) There is also the issue of whether the user took the advice because it was actually something needed or because it was presented by the smart and helpful agent (as was the case in Horwitz et al., 1998, mentioned in section 2.4.2).

The Othello game is a rather simple domain, and the combination of variables is small. In comparison, high functionality applications (MS-Office, Photoshop, etc.), serve a larger and more diverse user population. The design of these must adress problems such as (Fischer, 2001):

- the unused functionality must not get in the way;
- existing functionality which is unknown much be accessible when needed;
- it should not be too difficult to learn, use, and remember commonly used functions.

In these applications, users mostly want to get their task done - they do not feel a need to become experts. There is little to be gained for the user if a lot of effort has to be put into achieving only a small improvement. Fischer suggests that the payoff of cognitive artifacts can be characterized by the quotient of "value/effort".

Smart behavior of systems means that agents can guess wrong, sometimes performing changes that the user does not like. The lack of possibility or transparency for turning off the 'smart' features, means a loss of control for the user. On the other hand, the option of setting user preferences might be tedious (Dufresne and Hudon, 2002). The manner of portrayal of the agents presence presented in this paper, along with the user's possibility for control over it, is a possible solution for making interaction with the agent less annoying.

The well-contained domain of Othello was a graspable way of testing the suggested agent behavior of adapting to user feedback along with a more controllable graphic display of the amount of advice that the user can expect. It would be interesting to implement this in a more complex domain or as a part of a larger system. The user model might need to take a few more criteria into account, but simplicity might also lead to better transparency.

The aim of such an agent would be to offer more user control and a greater insight for the user as to how the agent is going to behave. In testing, how the user construes the agents behavior is central. Does the user feel that the agent responds to the user's actions in an appropriate way? Can the user influence the agent in an understandable and manageable way? Of course, the effect of the agent on the user's learning and other performance criteria is of importance, but if the user is to perform anything s/he has to be comfortable with the agent.

Appendix A: Instructions for test participants

All texts in the appendix have been translated from Swedish.

6.1 Instructions for playing with the static version

Your task is to play Othello against an invisible computer opponent. You play black and a game piece is placed on the board by clicking on the square where you want to put it. Your ability to play Othello will not be assessed.

An animated agent will, by pointing to a square on the game board, provide you with a suggested move every time it is your turn. It is only a suggestion and not necessarily the best move. You are in no way forced to take the advice.

6.2 Instructions for playing with the adaptive version

Instructions for playing with the adaptive version were the same as above, except that an extra paragraph was added:

The agent adjusts to whether or not you take the advice. If you take it, the agent will "take a step forward" and give you advice more often. If you don't take it, the agent will "take a step back" and give you advice less often. The agent moves up and down between two points. When the agent is furthest up/smallest, it does not give any advice, and when it is furthest down/largest, it always gives advice. Between these positions the amount of advice varies with the position: the further down the agent is, the more advice it gives. You can influence the agent's position more directly by placing the cursor on it and dragging it up and down.

Appendix B: Questions for test participants

The test participants answered questions one through eight after both versions, while questions nine through thirteen were answered after the version they played last. For the second set of questions, more space was left for filling in answers than is shown here. All texts in the appendix have been translated from Swedish.

Answer to questions 1-8 on a scale of 1 to 5, where 1 means not at all/never/very bad and 5 means a lot/always/very good.

1. How good would you say that the interaction between you and the agent was? 1 2 3 4 5
2. Did you think about why the agent suggested a certain move? 1 2 3 4 5
3. Did you wait for the agent's advice before you thought about what move to make? 1 2 3 4 5
4. Did you experience that the agent adapted to your behavior? 1 2 3 4 5
5. Did you find the agent to be distracting? 1 2 3 4 5
6. How much help was the agent's advice? 1 2 3 4 5
7. Did you experience that the agent behaved in a predictable way? 1 2 3 4 5
8. What was your overall impression of the agent? 1 2 3 4 5

   Answer to the questions below with your own words.
9. What did you think of the agent's appearance?
10. Did you think that the agent should have had more changes in facial expression? (For example, look happy when it gets to give a lot of advice and look sad when it does not.)
11. Have you used animated agents before this one? Where?
12. Was there anything that was worse in this agent? If yes, what?
13. Was there anything that was better in this agent? If yes, what?

References

Bates, J. (1994). The role of emotion in believable agents. In *Communications of the ACM*, 37(7), 122-125.

Baylor, A. L. (2000). Beyond butlers: Intelligent agents as mentors. In *Journal of Educational Computing Research*, 22 (4), 373-382.

Baylor, A. L. (2001). Permutations of control: Cognitive considerations for agent-based learning environments. In *Journal of Interactive Learning Research*. 12 (4), 403-425.

Baylor, A. L. (2002). Agent-based learning environments for investigating teaching and learning. In *Journal of Educational Computing Research*, 26 (3), 249-270.

Bickmore, T. and Cassell, J. (2000). "How about this weather?" Social dialog with embodied conversational agents. In *Proceedings of the American Association for Artificial Intelligence (AAAI) Fall Symposium on Narrative Intelligence*, (pp. 4-8). November 3-5, Cape Cod, MA.

Cassell, J. and Thórisson, K. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. In *Applied Artificial Intelligence*, 13, 519-538.

Chin, D. (2001) Empirical Evaluation of User Models and User-Adapted Systems. In *User Modeling and User-Adapted Interaction*, 11, 181-194.

Dautenhahn, K. (1997). Role of interactive conceptions of intelligence and life in cognitive technology. In J. Marsh, C. Nehaniv and B. Gorayska (eds.), *Proceedings of the Second International Cognitive Technology Conference*, CT'97, (pp. 33-43). August 25-28, 1997, Aizu, Japan, IEEE Computer Society Press.

Dautenhahn, K. (1998). The art of designing socially intelligent agents Science, fiction and the human in the loop. In *Applied Artificial Intelligence Journal*, 12 (7-8), 573-617.

Dehn, D. and van Mulken, S. (2000) Them impact of animated interface agents: a review of empirical research. In *Journal of Human-Computer Studies*, 52, 1-22.

Doyle, P. (1999): When is a Communicative Agent a Good Idea? In *Workshop on Communicative Agents of the Third International Conference on Autonomous Agents*, Seattle WA, May 1999.

Dufresne, A. and Hudon, M. (2002). Modeling the learner preferences for embodied agents: experimenting with the control of humor. In *ITS 2002 Workshop on Individual and group modeling methods that help learners understand themselves*. San Sebastian, Spain.

Duran, R. L. (1992). Communicative adaptability: A review of conceptualization and measurement. In *Communication Quarterly*, 40, 253-268.

Elliott, C., Rickel, J. and Lester, J. C. (1997). Integrating affective computing into animated tutoring agents. In *Proceedings of the IJCAI-97 Workshop on Animated Interface Agents: Making them Intelligent*. (pp. 113-121). Nagoya, Japan, 25 August.

Erickson, T. (1997). Designing agents as if people mattered. In J. M. Bradshaw (Ed.), *Software Agents* (pp. 79-96). Menlo Park, CA: MIT Press.

Etzioni, O. and Weld, D. (1995). Intelligent agents on the Internet: Fact Fiction and Forecast. In *IEEE Expert*, August, 44-49.

Fischer, G. (2001). User modeling in human-computer interaction. In *User Modeling and User-Apapted Interaction*, 11, 65-86.

Franklin, S. and Graesser, A. (1996). Is it an agent, or just a program? A taxonomy for autonomous agent. In *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*.

Gong, L. (2002). Towards a theory of social intelligence for interface agents. In *Human Factors 2002 Workshop Virtual Conversational Characters: Applications, Methods, and Research Challenges*. Melbourne, Australia, November 29, 2002.

Horvitz, E., Breese, J., Heckerman, D., Hove, D. and Rommelse, K. (1998). The Lumire Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI, July 1998, 256-265.

Höök, K., Persson, P. and Sjölinder, M. (2000). Evaluating users' experience of a character-enhanced information space. In *AI Communications*, 13, 195-212.

Isbister, K. and Doyle, P. (2002). Design and Evaluation of Embodied Conversational Agents: A Proposed Taxonomy. In *Proceedings of the AAMAS '02 Workshop on Embodied Conversational Agents - let's specify and evaluate them!'*, Bologna, Italy, July 2002.

Kaebling, L. P. and Moore, A. W. (1996). Reinforcement Learning: A Survey. In *Journal of Artificial Intelligence Research*, 4, 237-285.

Kobsa, A. (2001). Generic user modeling systems. In *User Modeling and User-Adapted Interaction*, 11, 49-63.

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. Chicago: The University of Chicago Press.

Laurel, B. (1997). Interface agents: Metaphors with character. In J. M. Bradshaw (Ed.), *Software Agents* (pp. 67-78). Menlo Park, CA: MIT Press.

Lester, J. C., Callaway, C., Grgoire J., Stelling, G., Towns, S., and Zettlemoyer, L. (2001). Animated Pedagogical Agents in Knowledge-Based Learning Environments. In K.D. Forbus and P.J. Feltovich (Eds.), *Smart Machines in Educations* (pp. 269-298). Cambridge, MA: MIT Press.

Lester, J. C., Convers, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI 97*, 359-366.

Maes, P. (1997). Agents that reduce work and information overload. In J. M. Bradshaw (Ed.), *Software Agents* (pp. 145-164). Menlo Park, CA: MIT Press.

Malone, T., Lai, K.-Y. and Grant, K. (1997). Agents for Information Sharing and Coordination: A History and Some Reflections. In J. M. Bradshaw (Ed.), *Software Agents* (pp. 109-143). Menlo Park, CA: MIT Press.

Middleton, S. (2001). Interface agents: A review of the field. Technical Report, University of Southampton.

Negroponte, N. (1997). From direct manipulation to delegation. In J. M. Bradshaw (Ed.), *Software Agents* (pp. 57-66). Menlo Park, CA: MIT Press.

Norman, D. A. (1997). How might people interact with agents. In J. M. Bradshaw (Ed.), *Software Agents* (pp. 49-56). Menlo Park, CA: MIT Press.

Porter, T. and Susman, G. (2000). Creating lifelike characters in pixar movies. In *Communications of the ACM*. 43 (1). 25-29.

Reeves, B. and Nass, C. (1996). *The Media Equation: How People Treat Computers, Televisions, and New Media Like Real People and Places*. New York: Cambridge University Press.

Rist, T., André , E., and Müller, J. (1997). Adding animated presentation agents to the interface. In J. Moore, E. Edmonds and A. Puerta (Eds.), *Proceedings of the International Conference on Intelligent User Interfaces* Orlando, FL. (pp. 79-86). New York: ACM Press.

Russell, S. and Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice-Hall.

Ruttkay, Zs., Dormann, C., and Noot, H. (2002). Evaluating ECAs What and how? In *Proceedings of the AAMAS '02 Workshop on 'Embodied conversational agents - let's specify and evaluate them!'*, Bologna, Italy, July 2002.

Shneiderman, B. and Maes, P. (1997). Direct manipulation vs.interface agents. In *Interactions*, 4, 42-61.

Smither, R. (1998). *The Psychology of Work and Human Performance*. Third Edition. New York: Longman.

Virvou, M. and Kabassi, K. (2002). Improving agent control for user modeling. In *First International IEEE Symposium on Intelligent Systems*, 73-78.

Walker, J. H., Sproull, L. and Subramani, R. (1994). Using a human face in an interface. In B. Adelson, S. Dumais and J. Olson (Eds.), *Human Factors in Computing Systems: CHI+94 Conference Proceedings*, (pp. 85- 91). New York: ACM Press.

Webb, G., Pazzani, M., and Billsus, D. (2001). Machine learning for user modeling. In *User Modeling and User-Adapted Interaction*, 11, 19-29.

Xiao, J., Stasko, J., and Catrambone, R. (2002). Embodied conversational agents as a UI paradigm: A framework for evaluation. In *Proceedings of the AAMAS '02 Workshop on Embodied Conversational Agents - let's specify and evaluate them!'*, Bologna, Italy, July 2002.