Evaluating Users' Experience of a Character-Enhanced Information Space

Kristina Höök, Per Persson, Marie Sjölinder
SICS, Box 1263, 164 29 Kista, Sweden
http://www.sics.se/humle/projects/persona/web/index.html
{kia, perp, marie}@sics.se

Abstract

We created the characters Agneta & Frida with the intent to strengthen and encourage exploration of information spaces. In a follow-up study we tried to capture whether users found the characters believable, whether they raised affective responses in users, and whether they created a richer more narratively oriented experience of the space. In order to do so, we had to develop new criteria and methods of understanding users' conceptions and affective responses. We discuss the study in detail, as well as the general implications for how to perform user studies and design of character-enhanced systems.

1 Introduction

Computer systems and services are complex phenomena. There seems to be two major general ways in which users and designers understand and make sense of information technology. On the one hand, computers are considered to be *tools*. The direct manipulation interface was designed with the specific purpose of enhancing this view of the interface. On the other hand, computers can be treated as more or less independent actors and intentional beings. This approach has occurred in many forms and been given many names in research literature, such as interface agents, believable agents, synthetic characters (see e.g. Elliott, 1992, Maes, 1994), not to mention the robots and androids of Science Fiction literature (Asimov, 1950). The 'tool-view' and the 'being-view' is not restricted to information technology only, but they are two fundamental 'stances' towards reality in general (cf. Dennett's (1987) distinction between mechanical stance and intentional stance, and Bruner's (1986) views on sense-making through narratives as a basis for human cognition).

When making computers more autonomous and pro-active, for instance through introducing characters in the interface, new ways of experiencing the interaction with computers arise. Systems of this kind will begin to invoke cognitive, emotional, social, aesthetic and ethical reactions not seen with 'traditional' user interfaces. This sets demands for new kinds of user studies to aid design and further our understanding of how these experiences come about.

1.1 Why focus on characters in information space?

In the PERSONA project[1], we viewed human-computer interaction in terms of the user being in an *information space*. In this space – which is to be taken in its broadest possible sense – users perceive, understand and bring meaning to the syntactical representations of the inter-

[1] PERSONA: PERsonal and SOcial NAvigation, an EC-funded i3 project:
http://www.sics.se/humle/projects/persona/web/index.html

face, and then act and interact on the basis of this. We designated a term for the whole of this process: *navigation*. In order to understand and address problems of navigation in information space, we differentiated between *wayfinding* and *exploration* (Benyon and Höök, 1997). In wayfinding, users have a clear understanding of what they want to find or do and they are relatively certain that the particular piece of information (or activity) they are looking for, is actually present in the information space. In a wayfinding situation, additional information or general suggestions from the information space may become a disturbance. The presence of other users, visual signals, sounds, and other cues in the environment may 'get in the way'.

In *exploration*, on the other hand, people are not trying to get anywhere; they are not trying to find their way. Instead, users want to get an overview of the space, explore the identity and function of objects. In an explorative mode, there is time and curiosity enough to get 'distracted' by side-tracks, to explore things that are not relevant to the major purpose of the navigation. In exploration, it is the 'richness' of the environment that determines the course of action and provides the quality of the experience of the space, and users may be more open to signals and suggestions form the environment. In this mode, it is the overall *experience* that matters: does the space encourage curiosity, does it feel interesting or entertaining, does it provide for a socially rich environment, or does it appeal to emotions and aesthetic values (Munro et al., 1999).

Wayfinding and exploration are not distinct categories, but rather ends on a scale. Some sessions will land inbetween. Users will also shift between modes within one and the same session. Distractions in a moment of wayfinding, can in the next – more explorative – situation turn out to be something interesting and enriching. Since the shifts between wayfinding and exploration can be floating (for instance, in web browsing) and since they are indeed two quite diverse modes of navigation, designing support for both of them can, indeed, be problematic. The same goes for measurements, and determining the success of a system. In particular, it seems to be difficult to capture the features of exploration in terms of the traditional usability criteria such as efficiency, ease of learning, and task completion.

We created the characters Agneta & Frida with the intent to strengthen and encourage exploration. We wanted to investigate the process of *exploration*, how 'infotainment' works, in what way anthropomorphism affects users, and what it means to equip believable characters with some form of personality. Moreover, we hoped to develop new methods and criteria for measuring users' understanding of, and socio-emotional reactions to, believable characters. Besides describing the system and study in detail, the major purpose of this paper is to discuss what we mean by these dimensions and how to capture them in user studies.

2 Believable agents: some previous user studies

Interface characters have been much criticised and debated in the HCI community (Shneiderman, 1997, Lanier, 1996, Suchman, 1997). They are said to violate good usability principles, to obscure the line of responsibility between human and machine, and to confuse both designers' and users' understanding of the computer's abilities and inner models of events. The proponents, on the other hand, regard these parameters as opportunities rather than reasons to avoid characters in the interface (Höök, 2000, Waern and Höök, 2000). Regardless of viewpoint, we need to develop frameworks to better describe their effect. If we claim that interface characters will raise expectations and appeal to users' emotions, we need to study and describe these processes. If we claim that they contribute to the naturalness of the dialogue with users, we need to sort out what we mean by a natural dialogue and study it. If believable agents promote the intentional stance of users and trigger anthropomorphic and socio-emotional reactions, we need to understand what those are. What are the relevant factors and

how do we measure them empirically? Let us discuss some types of effect that earlier studies have focused on.

2.1 Motivational effects in exploration

A number of studies have examined the ways in which characters enhance engagement and encourages exploration of a given information space, mostly in relation to learning and creativity. Such *motivational effects* were studied by van Mulken and colleagues (1998), who compared two versions of the presentation system PPP Persona: one with and one without a character. The study showed effects neither on recall of the presentation, nor on how the presentation was understood (objective measurements). However, it revealed a positive effect on the subjective estimation of whether the explanation was difficult or not. Subjects experienced the explanation as simpler with the PPP Persona character than without it. van Mulken and colleagues named this 'the persona effect'.

Another similar study looked at the persona effect for 'Herman the Bug', a pedagogical agent that helps students to create an ecological micro world system with plants, light and air (Lester et al., 1997). Here five different clones of the agent were compared, and the study revealed a persona effect – a strong positive effect on the students' perception of their learning experience. The animated character also had an effect on learning.

In a study by Wright et al. (1998) a plain textual explanation of a medicine was compared to one with the same text but with an animated dragon illustrating the different threats to the blood system. Here a negative effect on how much was remembered afterwards appeared; the dragon disturbed subjects, rather than aided them.

These conflicting results (PPP Persona and Herman the Bug, versus the dragon studies), point to the need for a better understanding of the design of synthetic characters in order to make use of their potential to encourage learning and exploration, and at the same time avoid the scenario in which the character distracts and disturbs the learning process. This involves, we think, a better understanding of the features of and relationship between wayfinding and exploration activities.

As pointed out by Andrew Stern (Hayes-Roth et al. 1998) (designer of the Catz and Dogz system) the artistic design and practical understanding of the creating of synthetic characters is crucial in determining the success of a system. A similar point is made by Elliott and Brzeinski (1998) when they cite Lester et al. (1997):

> "Lester gives the examples of, on the one hand, a humorous, lifelike, joke-cracking, character that ultimately impedes problem solving through his distracting presence; and on the other, a dull assistant that always operates appropriately but yet fails to engage the student. When communications from an agent must be coordinated to be both engaging and purposeful issues in timing, and the multi-layering of actions arise."

2.2 Anthropomorphic effects and believability

Another effect of synthetic characters is the ways in which they tend to raise expectations of anthropomorphism of the system (Reeves and Nass, 1996). Such *anthropomorphic effects* seem to have many dimensions. On the one hand the user may expect the system to be intelligent and cognitively potent. Brennan and Ohaeri, (1994) showed that users talked more to the anthropomorphic interface. King and Ohya, (1995) showed that users attributed more intelligence to anthropomorphic interfaces. Koda and Maes, (1996) showed that realistic faces are liked and rated as more intelligent than abstract faces. Opponents of synthetic characters argue that raised anthropomorphic expectations may lead to frustration in the user when the system

cannot meet the expectations (Schneiderman, 1997). For instance, the presence of a talking face might influence the user to expect the system to posses natural language and dialogue competence, which no system of today can live up to. The general conclusion is that the more 'natural' the interface, the higher expectations on intelligence in the system.

Besides anthropomorphism, which refers to any human-like characteristics in an interface that makes the user think that the system has some form of intentionality and human-like capacity for reasoning, the issue of character *believability* is sometimes mentioned in the literature. Exactly what is meant by believability is somewhat unclear. Some use the concept to refer to the facial expressions and body language of the character – the idea is that the more human-like and naturalistic, the more believable. Our view is somewhat broader, including the personality and attitude of the character, perhaps less focused on the bodily expressions. To our knowledge there has been no studies focusing on believability, in this sense, in the field of interface characters.

A possible exception is a study by Elliott (1997), where it was investigated to which extent a computer-generated face with spoken output (and music) could express recognisable emotions. The generated face was compared to a human actor. Overall, subjects did significantly better at correctly matching videotapes of computer-generated presentations with the intended emotion scenarios than they did with videotapes of a human actor attempting to convey the same scenarios. But the characters emotional state is but one aspect of believability.

2.3 Affective reactions

Third, some studies have focussed on the *affective reactions*[2] to synthetic characters. In a study by Sproull et al, (1996), a career counselling program was studied. Two interfaces were used: one textual and one facial interface. It turned out that users were more aroused (less confident, less relaxed) with the face, presenting themselves in more positive light. There were also gender differences: women preferred text, and men preferred the face. Emotional effects are of course often based on the cognitive understanding the user has of a given character or system, which involves very complex processes indeed.

Before discussing how we decided to measure the effects of Agneta & Frida in comparison to the studies presented here, we need to describe the Agneta & Frida system.

3 Agneta and Frida

From the web, we collected about 40 actual sites about film production, representing small production companies and organisations as well as local film production collectives. The sites presented information about present productions, production financing, manuscript sales, marketing of films, actors, co-workers, profile of the company, premiers, etc. Some sites contained an extensive body of documents, whereas others were quite simple in structure. Some were professionally designed, others of a 'home-page' character. All sites were in English, except one, which was authored in Swedish. We created an index of the sites and removed all outgoing links.

To this collected information space, we added Agneta and Frida. These two animated female characters – mother and daughter – would sit on the desktop, watching the browser more or less like watching television. With the help from a professional graphic designer and the voices of two actors, we created a library of short audio-visual animations or 'films', which

[2] Affective computing (Picard, 1997) is a newly emerged research field aimed at evoking and recognising user emotions.

would be triggered by three sets of cues.

First, most comments and behaviours of Agneta and Frida were related to specific locations in the information space. Downloading a page, clicking a link, dragging the mouse over an image or playing a soundtrack would execute general everyday speculations to what something on a site could mean, what the purpose with the site was, or if the design was likeable or not (see Figure 1). Error messages and browser malfunctions would be critically remarked upon. We programmed all of these comments beforehand. There was thus no real intelligence in the system. Often the comments alluded to Agneta and Frida's everyday life ('That looks like uncle Harry! [laugh]) and thus provided the user with their 'back story'. Since we envisioned Agneta and Frida to be computer illiterates, their remarks about computer technology and its male dominance were fairly critical and often ironic in tone. In this way, we scripted the comments in a way to reflect Agneta and Frida's personalities, ideologies, morals and sense of humour. Since both of them are strong-willed individuals, they occasionally even got into verbal disputes. None of the comments were scripted to be 'helpful', but to invoke laughter and sometimes reflection on the information browsed. Although a user might come back to the same location, none of these behaviours were executed more than once.

INSERT FIGURE 1 ABOUT HERE

A second set of behaviours / comments was of a more general nature, unrelated to content or user's activity in the information space. This included blinking, picking noses, going to the toilet/kitchen, drinking coffee or general gossiping about uncle Harry and Miss Andersson (the owner of a repulsive poodle that often enters Agneta and Frida's back yard, and about which they occasionally fantasise killing). These would be triggered at certain intervals when there were no other behaviours running. This set of behaviours was included to create more lively characters, with lives independent of the happenings in the browsing session. We wanted to avoid the impression that the behaviours were only automatic reflexes of user's actions. Some behaviours in this category reoccurred now and again (e.g. blinking, yawning and sighing).

Thirdly, we allowed the user to choose among a small set of narratives that would run in parallel with the browsing session. By clicking on the right mouse button the user would be presented with a menu in which she could make the choice between different genres and titles. The system provided one comedy ("Poodles: Cute Fluff or Ambassadors of Evil?") and one melodrama. Each of these plots involved about ten scenes, which by certain time intervals were played out on the desktop, mingling with the other two kinds of behaviour. In contrast to the first set of triggered behaviours, the narratives were not related to the content of the web pages. In contrast to the gossiping behaviours, the narratives provided a sequence of events that was causally coherent, with a clear beginning and an end. (Only one of the subjects in the study chose the narrative mode.)

In all three behaviour types, Agneta and Frida spoke English with exaggerated Swedish accents (the Swedish Chef of the Muppet Show acted as our leading inspiration here). As we aimed for users between 15 and 30 years of age, the web page contents, the commentaries as well as the jokes of Agneta and Frida, were all scripted with that age group in mind.

The user could regulate the intensity of Agneta and Frida's behaviours. Through a pop up menu, the user could set the level of activity (0-5) which in different degrees disconnected some behaviour and defined the time interval between the remaining ones. The way to access this menu and the other features of the system were described or hinted at by Agneta or Frida

when they were clicked upon.

Another feature was a search engine situated in their living room drawer on the right hand side of the screen. Here the user was able to search not only for strings in the documents as in a normal search engine, but also for words in the Agneta and Frida comments history. If the search were successful, the user would be presented with the joke or comment in which the search string occurred, as well as the actual URL.

We implemented these ideas using JavaScript and Microsoft Agent Tool.

The fact that the user is confined to a selected set of web pages, and that the system does not provide any machine-generated jokes, comments or behaviours but simply 'films' that are being triggered at specific nodes in space, makes Agneta and Frida less of an actual artificial intelligence system, and more of a mock-up or test-bed for experimentation of design.

3.1 Design Rationales

The design of Agneta and Frida was guided by a number of rationales.

Exploration

We wanted to support exploration rather than wayfinding. We did not expect users to solve tasks in the information space, or to find a particular piece of information, but rather browse around and explore. Both the system – Agneta and Frida's commentaries were not trying to support serious information seeking – and the experiment situation (described below) were designed with this mind.

Affective reactions: Experience of space

We wanted to explore and evoke user emotions. Being in and navigating through information space is of course influenced by 'rational' decisions and reasoning, but this cognition is always accompanied with more subjective and emotional dimensions, particularly – we reasoned – in exploration. We wanted to trigger feelings related to likeability, identification, humour, 'situation', irony, frustration, anger, social emotions and moral judgements. We wanted to enrich the *experience* of information space. By designing the agents with a definite humorous twist, in many cases connected with the information displayed on the site, we wanted to enhance affective experiences of web browsing and information seeking.

Believability

We wanted to promote believability of the characters by giving the users cues to their background, personality and situational context. In contrast to other more anonymous embodied agents like the presentation agent of PPP Persona (see section 2.1), we attributed Agneta and Frida with a 'life': Frida is 30 years of age, still lives at home, has problems getting boyfriends, her father is absent, non-academic, works in an office, dislikes Uncle Harry. Agneta loves coffee, refuses to talk about Frida's father, hates poodles, is poor but strong-willed. Placing character in a setting like this, of course, goes hand in hand with the goal of promoting a narrative experience.

Agneta & Frida's body language is quite limited and non-expressive in itself. Instead we hoped that their inner lives, their attitudes, their humour, their voices, and the interaction between them and the user, would provide for their expressivity and believability.

Narrativity

We wanted to promote a narrative experience of the information space. Many users appear to experience hypermedia in terms of spatiality. Everyday speech reflects this stance: 'we browse/surf the web'; 'we go to pages'; 'we enter/leave pages'; 'pages contain information';

'the web is an information space in which we look for things' (cf. Maglio and Matlock, 1998). There also seems to be a correlation between cognitive spatial ability and how a user finds her way in hypermedia (Dahlbäck et al, 1996). We hoped that Agneta and Frida's comments, jokes and narratives would encourage the user to construct coherence between nodes on the bases of events, situations and psychology of the characters, rather than along spatial dimensions. We hoped to merge spatial and narrative modes of experience.

Irony and critique

We wanted the system to be 'critical' in several respects. First, the mere presence of Agneta and Frida challenges the standard view of computer technology as direct manipulation tools through which we accomplish 'serious' tasks and wayfinding problems. Second, Agneta and Frida's brutal and often sarcastic critique of web design and system failures provides a self-reflexive system, commenting and making fun of the conventions and short comings of computer technology. The facts that the characters are turned towards the browser and watch the information space *along* with the user emphasises this outsider view vis-à-vis the system. Thirdly, Agneta and Frida miss no chance to make visible and make fun of the male dominance in society, computer community, and in film production.

Two characters interacting

We wanted to explore the possibilities of multiple characters. By having two characters instead of just one, we made pre-scripted dialogue possible, which enabled us to introduce personality, humour, negotiation and self-reflection more naturally than through a single character. In fact, few systems exploit this design choice, with the exception of (André and Rist, 2000).

Off-screen desktop-space

Finally, we wanted to expand the desktop metaphor. On the one hand, we placed Agneta and Frida in a living room environment (with chairs and a drawer). On the other, we expanded the on-screen space into the non-visible areas. Analogous to cinema and VR environments (see Persson, 1999), the Agneta and Frida system constantly invokes and plays with this *off-screen space*: off-screen left is a kitchen and toilet, which Frida occasionally visits (audible on the soundtrack). Off-screen right is the back door and a yard, which often is visited by Miss Andersson's poodle. With an active employment of sound techniques, we tried to encourage the user to re-conceptualize the notion of a 'desktop'.

In the ensuing study, we focused on dimensions of emotions / experience, believability, exploration and narrativity, thus leaving aside users' understanding or liking of the irony, the effect of having two characters versus only one, and effects of the usage of off-screen space.

4 What to measure?

In designing the study of Agneta and Frida, we reviewed the earlier studies and found some shortcomings that we wanted to address in our study. Let us discuss each measurement we took, and our reasons for choosing them.

4.1 Exploration

We decided to investigate whether the system encourages exploration through comparing using Agneta & Frida with just simply navigating the web sites without the presence of them. We measured exploration through questions of whether subjects thought the characters were fun; if they wanted to return to the sites; if they paid attention to text or images commented upon by Agneta and Frida.

We also asked users how long time they thought that they spent and compared that time to actual time spent. Our hope was that if Agneta & Frida encouraged exploration users would perceive it as if 'time flew' away. We asked them how many pages in the web site they visited. We also interviewed users and gave them a questionnaire.

Even if we did not expect Agneta & Frida to contribute to any other more objective measures of efficiency, we did measure to what extent subjects remembered web pages better if they had heard Agneta & Frida make jokes about them.

4.2 Affective reactions

Second we wanted to measure the users' emotional reactions vis-à-vis Agneta and Frida. From the area of affective computing (Picard, 1997) we knew that bodily reactions might be a means to understand users' affective responses. Picard and colleagues measure the physical responses of the user through EKG, skin conductance, blood pressure, and facial expressions. Such physical measurements may provide some quantitative evidence of users' experience of interfaces, but the relation between a feeling and its physical expression is far from direct and unproblematic. First, people differ radically in respect to their body language. Some people make constant use of gestures, facial expressions and bodily reactions. Others are more subdued and restrained. Second, bodily reactions can be caused by many different things. The 'meaning' of a smile, for instance, may shift abruptly between different situations (the person thinks something is funny, the person is embarrassed, the person wants to hide some other emotion, the person smiles because other people smile etc). Third, EKG and skin conductance may measure general agitation, but they will not provide any nuanced analysis of the specific emotions behind this agitation. Such methods, are blunt instruments in measuring emotions, unless combined with an interpretation of the situation context (see e.g. Paiva and Martinho, 1999).

Thus, if we aim for a qualitative understanding of user's emotions, physical measurements on their own are not sufficient. Such methods have to be complemented with some understanding of the context in which these reactions arise. This involves the user's cognitive, emotional and personality preferences, body language, sense of humour, or the situation in which the expression is taking place.

In a short study like this, however, it is impossible to measure all of those aspects. We decided on a combination of physical and qualitative measurements. We videotaped user's face during the use of Agneta and Frida, and counted the number of smiles, giggles and laughs. We also measured the time spent and the user's subjective estimations of time spent. Since the task was explorative in nature, user's could stay in the information space as long as they wanted, we assumed the absolute time would reflect some aspects of emotion: if the system triggered positive emotions, users would stay longer. If the system triggered negative emotions or no emotions at all, people would stay shorter. In addition to this, we asked about their mood before and after and using the system. We also made interviews and questionnaires dealing with their emotional experience of Agneta and Frida. Our intention was to capture some of subjects' attitudes towards and interpretation of the system.

4.3 Anthropomorphism and Believability

Anthropomorphism is often talked about in terms of intelligence. In the studies cited in section 2, interface characters seem to trigger higher expectations of intelligence than non-character interfaces. Although we do not disagree with these results, we do think that the notions of anthropomorphism and believability have to be expanded to include other types of expectations. Our basic assumption here is that characters – whether they appear in comput-

ers, fiction or cinema – are not only made sense of through the features of face and body, but through the ways in which character act within situations, display humour and share values and moral perspective of the user (Tan, 1996). It is, for instance, not enough to place characters in a film in order to evoke emotions (and thereby commercial success). These characters must move, act, feel, think and talk in concrete situations in ways that raise the spectator's sympathy or antipathy. As pointed out by Porter and Susman (2000) when discussing how to create life-like characters in cartoons ('Toy Story'):

> "They [animators] understand that 'life-like' does not mean 'has movement'; lifelike means 'has a brain'. The underlying notion of Pixar and Disney animation is that action is driven by the character's cognitive processes – that it reflects intelligence, personality and emotions."

These dimensions, however, are much more difficult to design for and evaluate.

Concerning the anthropomorphic expectations we put several statements and questions in the questionnaire, e.g.:

Questions with a scale from 1 – 7	**Statements to agree/disagree with**
Was it possible to predict what Agneta & Frida would do next?	I think that Agneta & Frida would enjoy having me together with them in front of the browser
Did the animated characters feel 'human-like'?	I know someone who is like one of the characters
Did the animated characters feel like friends?	Agneta & Frida seems to know as little as I do about independent movies
	Agneta & Frida would like my home page
	Agneta & Frida do not seem to like men

To capture issues of believability, we also asked subjects to freely describe their experience of Agneta & Frida, as if they were describing it to a friend. From these interviews we then collected statements where they talked about Agneta & Frida in an anthropomorphic way.

4.4 Narrative Experience

We wanted to create a more narrative experience with Agneta & Frida. We often think about the world in spatial terms, which is not very surprising: our bodies are symmetrical, there is a behind and in front of, above and below, an inside and an outside, we move through a spatial world. This colours the way we speak about the world, as is apparent from the metaphors we use (Lakoff and Johnson, 1980; Lakoff and Johnson, 1999). One of the functions of metaphor is that it helps people think about relatively abstract conceptual domains in terms of relatively concrete domains. From the work by Maglio and Matlock (1999) we knew that web browsing is often perceived as a spatial activity: the user is viewed as an agent moving through the space of sites and web pages. Maglio and Matlock found this through examining what metaphors were used when subjects described their surfing through web pages.

With Agneta & Frida, we hoped to merge spatial and narrative modes of experience. We therefore decided to analyse the language used in the interviews with the subjects from a metaphorical point of view. There are many different methods we could use to do this. One is to go through each sentence and try to do a semantic analysis to determine which metaphor had influence the description. Another is to take a number of measures, such as the number of verbs, adverbs, pronouns, etc. of the two categories (narrative and spatial). The latter is a more blunt tool, but one that is less dependant on a subjective analysis. It will only be able to say

something about the likeliness that a certain metaphor was used more than another. We decided to go for this approach focusing on the verbs and adverbs in the interviews – which also follows the method used by Maglio and Matlock (1999). Spatial verbs and adverbs were characterised by movement (e.g. 'going through'). As for narrativity we looked for words containing temporal dimensions (e.g. '…and then…') or intentional/psychological words (e.g. 'giving up'; 'bored'; 'anxious'). In contrast to spatial experiences, narratives are temporal chains of events, which are experienced as causally related. Most often, this causality often involves characters and their mental states of different sorts (beliefs, emotions, perceptions etc.). We assumed that temporal and psychological terms would reflect this kind of experience. This, of course, is not in contrast to perceiving and talking about the world in spatial metaphors. A story may well contain spatial metaphors, spatial relationships between event, people and objects. Narrative descriptions (or metaphors) add the relationships between events and often contain characters that act because of their inner emotional states.

Given these quite complex and ambitious goals of our user study, let us now turn to the actual study. Afterwards we shall discuss the limitations of the study, and the implications for future studies of believable characters.

5 METHOD

Thus, there were four questions that we focused on:

- Does Agneta & Frida encourage users to explore the space?
- Does Agneta & Frida create affective responses in users?
- Are Agneta & Frida perceived as believable characters?
- Does Agneta & Frida provide for a more narrative experience?

5.1 Subjects

The 38 subjects were recruited to be between 20 and 30 years. 18 subjects used Agneta & Frida (the 'withA&F' group) and 20 subjects explored the web sites without the characters present (the 'withoutA&F' group).

The subjects of the withA&F group were in the range $19 < 26{,}2 < 41$ years old. 7 women and 11 men. 10 had a technical background the rest had other professions. All but one had a university degree. The subjects in withoutA&F group were in the range $22 < 26{,}6 < 32$ years old. 12 women and 8 men. 14 had a technical background. All but one had a university degree. All had a good understanding of English.

Subjects were signed up through friends and colleagues, and were given movies tickets in return for participation.

5.2 Tasks and procedure

The withA&F group was first asked about their age, gender, occupation, education, English-, computer-, and web experience, as well as their present mood – scaled from 1 (lousy mood) to 7 (excellent mood). Wrist watches were removed so that they would not keep track of time.

Subjects in the withA&F group got the following instruction: "Imagine this situation: you are at home one evening, and you have nothing in particular to do and you are not especially tired. You have a fast and efficient computer at home with a good and fast connection to the Internet. A friend has suggested some cool web sites on the net that you should have a look at. Check out the web link the same way that you would do if you were sitting at home. I'll be in my room over here, and you can come and get me when you are finished". The experimental

leader would at this point leave the room to allow subjects to feel free to do what they wanted and stay as long as they wanted. The system was started, and Agneta & Frida would appear on the screen. Subjects' interactions with the system and facial expressions were video-recorded.

Once they were finished, they were asked to grade their mood after using the system on a scale from 1 (worse) to 7 (better). They were then interviewed about their experience, based on the question: "If you met a friend downtown and was asked to describe what this system was all about, what would you (in as many details as you can) say?". Interviews were taped on audio-cassette.

They filled in a questionnaire[3] with three sets of questions:

- Estimated time spent and estimated number of pages visited.
- A number of Likert-scale questions on whether they perceived Agneta & Frida as believable characters and whether they found them entertaining or disturbing. In total, 14 questions were asked in various ways in order to get at users' experience of the system.
- Finally, they marked which of a set of 20 statements they could agree with. There were statements such as "I've get the same kind of humour as Agneta & Frida", or "It felt good to have two ladies to browse with".

Once the questionnaire was completed, they were shown 38 screen-shots of web pages from the sites and were asked if they had seen them and whether they remembered any jokes Agneta & Frida had made about that web page.

Finally, subjects in the withA&F group were asked to freely comment the system.

In total, it took each subject about 2 hours to complete the steps. Afterwards we explained the study and the design rationale behind Agneta & Frida. Finally subjects were given the movie tickets.

On the basis of the video-recordings, we estimated how long time they actually spent with the system. This time was compared to how long time they assumed that they had spent. The video-recordings of the facial expressions were analysed. Here we counted smiles – smiles proper, giggles and outright laughs. We constructed a variable – smiles per 10 minutes – consisting of the total number of smiles, subtracting the number of negative smiles (laughing *at* Agneta & Frida) and unrelated smiles (arising from the study situation or the web page content *per se*). The experimenters made this categorisation on an intuitive basis. The number of smiles was supposed to capture some of the non-conscious affective reactions.

The withoutA&F group went through the same steps as the withA&F group, except all questions related to Agneta & Frida were removed from the questionnaire, and their facial expressions were not recorded.

6 Results

We divide the results of the study into the four questions from above.

6.1 Does Agneta & Frida encourage users to explore the space?

The withA&F group claimed that they visited in average 69.3 pages, while the withoutA&F group claimed to have only visited 45.0 pages in average. We did not count the actual number of pages visited. As we noted above, the withA&F group spent more time with the system (27.4 minutes/subject versus 20.7 minutes/subject). The longest time spent was 70 minutes for

[3] The full questionnaire (in Swedish) can be found at: http://www.sics.se/~kia/questionnaire.htm

the withA&F group and 42 minutes for the withoutA&F group. The shortest time spent was 6 minutes for the withA&F group and 3 minutes for the withoutA&F group. It seems as if Agneta & Frida made subjects stay longer in the space and explore more pages.

There was no difference between the two groups in terms of how much they remembered of the space. Out of the 38 randomly selected test pages, the withA&F group remembered 88% of the pages they had seen, while the withoutA&F remembered 89%. There was no difference between pages with or without comments from Agneta & Frida. Notable was that subjects were able to accurately recall the comments Agneta & Frida had made at particular pages.

We noted that several users claimed that they felt relaxed when navigating together with Agneta & Frida. 14 out of 18 subjects agreed with the statement "It felt good to have two ladies to browse with". Some users on the other hand, got very irritated and felt that Agneta and Frida got in their way. Out of 18 subjects, 4 were often disturbed, 6 were sometimes disturbed, while 8 were never or almost never disturbed. 6 subjects found that Agneta and Frida often stole attention from the information in the web pages.

We found a correlation between how much subjects were disturbed by Agneta & Frida and their web- and computer experience. Users who had a lot of web experience were also more disturbed by Agneta & Frida ($r=.54$, $P < .05$), the same for computer experience ($r=.60$, $P < .05$).

In summary, while Agneta & Frida did encourage exploration, they also disturbed some users more than others (cf. the discussion section 1.1).

6.2 Does Agneta & Frida create affective responses in users?

One might think that it is fairly straightforward to divide the group of subjects exposed to Agneta & Frida into two groups: one that liked them and one that did not. But when comparing the following time spent, mood after using the system, questionnaire-replies, and facial expressions, we arrive at a more complex picture.

Time spent versus perceived time

Time measurement showed that the subjects in the withA&F group spent in average more time than the withoutA&F group (27.4 minutes/subject versus 20.7 minutes/subject). This might indicate that the experience with Agneta & Frida was richer, more interesting, and therefore made users stay longer. This result was consistent with the second time measurement: the subjects in the withA&F group *under*estimated time by 2.5 minutes in average, while the withoutA&F group *over*estimated time spent by 3.1 minutes. This result seems to be in line with our expectation that Agneta & Frida creates a richer experience. The difference in estimated time is however, too small to judge.

Interestingly enough, users that smiled a lot, kept on smiling even after an hour of use.

Questionnaire replies and number of smiles

While the time measurements were encouraging, the questionnaire replies complicated things. As seen in table 1 and 2, there was a correlation between the different judgements from the questionnaire (for the four questions related to their experience of Agneta & Frida) in the withA&F group, which was expected. However, as table 1 also shows, the number of positive smiles per 10 minutes did not vary with users' judgements (no significant correlation). That is, subjects who claimed to like Agneta & Frida did not necessarily smile a lot (as for e.g. subjects 3, 4, and 6), while those who smile a lot sometimes claim not like them at all (as for e.g. subjects 12, 16, 17).

INSERT TABLE 1 AND 2 ABOUT HERE

Furthermore, the measurement of the difference between estimated time and actual time spent followed neither the questionnaire replies nor the smile measurement (again a non-significant correlation). This was also the case in respect to the absolute time measurement (a non-significant correlation value). Subject 16, for instance, smiled as often as 7.5 times per 10 minutes, spent 36 minutes (9 minutes above average) with the system, which would indicate that he had a good time. However, his post-usage view on Agneta & Frida's commentaries was only 3 on the 7-grade scale. On the other hand, subject 1, who smiled the least, only 1.2 smiles per 10 minute, and only spent 16.5 minutes with the system, really liked Agneta & Frida – giving them grade 6 on the 7-grade scale.

Mood

The mood measurement, depicted in table 3, shows that the withA&F group claimed to be in the same or better mood afterwards, while the withoutA&F group was in the same or worse mood. This indicates that Agneta & Frida's humorous commentaries and behaviour did, for most subjects, provide a positive experience.

INSERT TABLE 3 ABOUT HERE

But again, the change in mood did not follow any of the other measurements, see table 4 (no significant correlation).

In summary, neither of the five different measurements (estimated time spent, difference between actual and estimated time, number of smiles per 10 minutes, questionnaire replies, and mood after using the system) were correlated.

INSERT TABLE 4 ABOUT HERE

Angry with Agneta & Frida?

Despite the fact that most users were in a better mood after using Agneta & Frida, several users said that they at several points were angry with Agneta and Frida. In fact, a couple of users were angry with them all through the experiment (see table 5). One subject said (translated from Swedish):

> ”The animated characters were extremely disturbing and distracted me completely with the effect that I could not concentrate on the browsing and lost interest in the contents of the web pages. The surfing quickly turned into fear of what Agneta & Frida would say next.”

On the other hand, some subjects really liked them and were neither disturbed nor angry with them. One said:

> ”The ladies enlighten the atmosphere. Nice with some company so one does not become completely stiff and dry.”

Not everyone liked them, or found them funny, but nobody was indifferent to them.

INSERT TABLE 5 ABOUT HERE

6.3 Are Agneta & Frida perceived as believable characters?

On the question "Did the cartoon characters feel 'human'?" most subjects answered that they were, see table 6. Subjects who found the characters fun and nice were more inclined to ascribe them human characteristics ($r=.41$, $P < .05$). Subjects that found the characters fun and nice also, not surprisingly, wanted the characters to act more and were more willing to use the system again ($r=.52$, $P < .05$). But not everyone felt as if Agneta & Frida were their friends, see table 6. We might very well understand and find a character believable, but that does not mean that we like them or would like to have them as friends.

INSERT TABLE 6 ABOUT HERE

Subjects did not feel that they could predict what Agneta & Frida would say next (on the 7-grade scale, subjects averaged on 5.55 where 7 meant 'not predictable at all'). This means that the characters were not perceived mechanical and machine-like, but had varied comments. On the other hand, one might claim that a believable character should be somewhat predictable – that it should be possible to get to know the character and thereby start anticipating the character's responses while still avoiding a too stereotypical, one-dimensional character.

On the statements that subjects could agree or disagree with, we got the following results:

- "I know someone who is like one of the characters" 9 subjects agreed
- "Agneta & Frida seems to know as little as I do about independent movies" 14 subjects agreed
- "I think that Agneta & Frida would enjoy having me together with them in front of the browser" 8 subjects agreed
- "I have the same kind of humour as Agneta & Frida" 9 subjects agreed
- "Agneta & Frida would like my home page" 1 subject agreed
- "Agneta & Frida do not seem to like men" 2 subjects agreed

In the interviews, the subjects talked about the characters in an anthropomorphic way. They said that the characters made them feel less lonely or that the characters were enervating. Other subjects' comments concerned the characters behaviour or feelings, for instance "it is fun that the characters are far too self-assured" or "the characters thought it was dull when I did nothing for a while…"

Several subjects also wanted to be able to answer, interact with, or even get even with the characters when their comments had been nasty.

Comments made in the interviews made it clear that subjects had understood that Agneta & Frida are mother and daughter, that they are poor, which explains why they sit in front of the web browser instead of buying a TV, and that Frida is not married, and she is more sarcastic and often explains stuff to Agneta. For example, one user said (translated from Swedish):

> "Yes, they are relatives and they do not really get along that well but they are stuck, they cannot do anything else. If they had the possibility to do something else, they would. Then… Frida is the daughter, right? She started looking at me, she turned around and looked at me when I did something that she did not like, or, I do not know the intention of that… […] so they seem to be tired of what they are doing".

The users understood both the social aspects of the characters and the type of situation these acted within (watching TV/browser in the living room).

6.4 Does Agneta & Frida provide for a more narrative experience?

As discussed above, we tried to see whether subjects talked about their experience of the system in more spatial or narrative terms. We did this through counting the number of verbs and adverbs that can be said to be more spatially versus more narratively oriented. An example of a spatial coding of one of the interviews (translated from Swedish):

”**I have surfed on** the Internet and I **came into** a site that dealt with independent movies and there were all sorts of weird places. Oh, and among other things something from Canada, but unfortunately one had to download so much **there** so I gave up that page and **went away from there**. Eh... then there were some different eh film companies and such.”

An example of narrative coding of the interviews:

”I have surfed on the Internet and I came into a site that **dealt with** independent movies and there were all sorts of weird places. Oh, and among other things something from Canada, but unfortunately one had to download so much in there so **I gave up** that page and went away from there. Eh... **then** there were some different eh film companies and such.”

When looking at the metaphors used by subjects in the interview after using the system, we found that the withA&F group talked about their experience to a larger extent in terms of narrative verbs and adverbs (68% narrative), while the withoutA&F group used more spatial verbs and adverbs (only 45% narrative), (Tables 7[4] and 8). The difference between the conditions is statistically significant (Mann Whitney[5]: $p>0.95$). The comparison was made by comparing the number of spatial expressions and the number of narrative expressions, normalised by the number of analysed verbs and adverbs.

INSERT TABLE 7 AND 8 ABOUT HERE

While this analysis seems to support our hope that Agneta & Frida did indeed created a narrative experience, a careful reading through the interviews revealed that subjects had not blended the two experiences. In the interviews, they separated the description of the web sites from their description of Agneta & Frida. The first would be done through using numerous spatial metaphors, while when they talked about Agneta & Frida, they used a lot of affective, pronominal, and narrative expressions.

7 Discussion of results

The methods and results of this study raise interesting questions and lessons to be learned. The study exemplifies the difficulties in measuring experience of an infotainment application – or to be more precise, what measurements more truly reflect subjects’ overall experience of the system.

Our study showed that the Agneta & Frida system encouraged subjects to explore more of the

[4] The classification of which verbs and adverbs can be considered to be spatial follows the classification by Maglio and Matlock (1999).

[5] Mann-Whitney was used since usage of spatial versus narrative verbs and adverbs cannot be assumed to be normally distributed.

space and to stay longer in the space. But subjects did not learn more about the space, and some got distracted and disturbed by Agneta & Frida. The interesting result here was that computer- and web-experienced users were more disturbed than others were. We can only speculate about the reasons for this. Maybe they have a better-defined strategy for how to navigate the web, mostly directed at wayfinding problems, and anything that comes in the way is perceived as disturbing?

The study also shows that we succeeded in our attempt to create believable characters with a life, personality, attitude, and humour. Half of the subjects agreed with statements that indicated that they had the same humour as Agneta & Frida – and humour, we believe, can only be ascribed to believable characters. That is, if we take believable to mean something more than 'has movement'.

Our metaphor-analysis revealed that Agneta & Frida provided for a more narrative experience, but our intention to make subjects melt the whole exploration session into one big story was not met. The two parallel stories, the web page content and the commentaries of Agneta & Frida, are probably each too strong and not enough related to make it possible for users to experience the whole interaction as one coherent story. With a more careful scripting, perhaps in an educational setting, we believe that it would be possible to create more of a narrative coherence.

We did not allow for many interactions between the user and the characters. The user drove the web browsing which in turn evoked responses from Agneta & Frida. Enabling a richer dialogue interaction between the characters and the user would be an interesting extension. Such a dialogue would allow us to explore *interactive* narratives (Murray, 1997).

Finally, we come to the fact that subjects' facial expressions or time spent in the system did not predict their questionnaire-evaluations. How are we to explain the fact that none of the five variables discussed in section 6.2 correlated?

Fuzzy measurements?

One way of accounting for the difference between the measurements is to claim that each of them is afflicted with problems and errors. Counting the number of smiles, for instance, is problematic since people may have different body language. The presence of a video-camera might make some people smile more, others less. When analysing the videotapes, we noticed that some users were more inclined to ignore the audio-based comments from Agneta & Frida, and instead focus on the content of the web pages, whilst others were more alert and attuned to the jokes. This affects the degree to which people smile. In addition, smiles can mean different things: one may, for instance, smile because one likes something, because one finds it silly, or as a reflection of someone else's smile or laughter. Although we tried to filter out these variables, they still might have influenced the data.

Likewise, the results from Likert-scale question as "Agneta & Frida's commentaries are fun", might be influenced by social processes of the experiment set-up. People are generally known to behave in a socially desirable way, i.e. according to what they believe the experimenter desires. Since experimenter and designer in our study could be considered to be one and the same by some subjects, it is likely that such 'politeness' effects were present.

But even if the measurements are fuzzy, the fact remains that the five different measurements do not follow one-another in any consistent way.

Measuring different aspects of experience?

Another way of explaining the non-correlation, is to assume that the variables simply measure

different things. We believe that although all of them try to capture the overall experience of the system, they may, in fact, measure different *aspects* of this experience.

For example, facial expressions may provide indications of the immediate appreciation of the system. Facial expression will perhaps show the instantaneous reactions to the jokes, but not the retrospect evaluations of Agneta and Frida. The post-usage replies, on the other hand, might reflect subject's 'after-thoughts' about the system, which may be influenced by moral and ethical preferences. Such retrospect judgements may be influenced by moral, ethics, and more official views of what humour and entertainment should or should not be. For instance, one might first laugh and then still afterwards reconsider and review the joke as 'vulgar', 'silly' or 'representing bad taste'. The post-usage evaluations may reflect more authorised views on humour, interfaces, and interface characters. In this respect we found the correlation between computer experience and the whether subjects were disturbed by Agneta & Frida elucidating: computer experienced users may have a strict and 'Lutheran' model of computer interfaces and web browsing. Since Agneta and Frida blatantly break with this 'tradition', experienced users are more disturbed than users who do not have such strong expectations or 'preconceptions'. Especially subjects who are used to having complete control over the computer – from the insides of the operating system and out – may find it hard to accept characters in the interface and processes that run outside their control. In fact, before, after, or even during the session, some subjects said that they in general disliked interface characters.

The mood measurement – which lands somewhere in-between the instantaneous reactions during use and the post-usage replies – will again measure something else. Since it showed that the Agneta & Frida subjects were in a better mood compared to the without Agneta & Frida subjects, it provides us with some evidence that the system positively influences users' experience of the system on an emotional level.

Experiencing a given system involves many aspects and their complex relations require a careful analysis of measurements taken. In upcoming studies we have to consider these parameters more thoroughly.

What aspect of experience is most important – and thus determining the appropriate method of measurement – is of course dependent on the design goals. If we aim to entertain for a one-time usage situation, then maybe it is more important that subjects smile a lot; if we want subjects to return to the space, then their post-usage evaluation should be emphasised. The fact that many users were disturbed by Agneta & Frida – but still enjoyed their company – indicated that we failed to create a feeling of flow or relaxed relationship to the space. If that is our design goal, than other means should be used.

Providing a richer context for interpreting results

On another level, our results also point at the difficulty of gathering facial expressions and using those as a means to measure subjects' affective reactions towards computer systems. As discussed in section 4.2, users' physical reactions of interactions with systems are not necessarily good predictors of users' inner mental states. It is possible to see that the user gets aroused, but not whether it is a positive arousal or not, nor whether it reflects a mental state that the system should adapt to (Riseberg et al. 1998). In order to pinpoint finer distinctions in the emotional reactions, we have to consider the users interpretation, understanding, attitudes, and expectations of computer culture. The experience of jokes and irony, for instance, will be determined by personal expectations, but also by social and cultural context. Our views on humour are reflections of our personality and who we want to be in the eyes of others. Sometimes Agneta & Frida make strongly ironic and sarcastic remarks about the computer and web culture:

Frida: They say that computers save so much time. But sometimes I wonder… At work I often feel like I'm spending 90% of the time getting the damned thing to work, and about 10% of the time actually accomplishing things with it….

Agneta: I don't really know… I'm not that experienced…

Frida: Maybe we should buy a home computer…? Just for the fun of it…

Agneta: Naa, I'd prefer a television set instead… there are more stories on TV….

Some jokes are concerned with the male dominance of the IT-world:

Frida: Stupid! Nothing works! Who would ever publish a page like this?

Agneta: A man?

Users might approve or disapprove of this type of humour or the views of Agneta & Frida. In order to determine and predict such processes, we would need a thorough investigation of subjects' attitudes towards humour, irony and fictional characters in general and attitudes towards these phenomena in interfaces in particular. For instance, we should have asked subjects about their general attitudes towards interface characters in order to better understand the positive and negative reactions Agneta and Frida triggered. Since the system breaks with traditional GUI interfaces (and tries to support non-wayfinding activities), such attitudes of users are important factors. Although some parts of the questionnaire and the interviews indicated such background preferences of the users, we feel that we should have put more emphasis on these parameters. If we had known more about the subjects we would perhaps have a better understanding of the non-correlation between facial expressions, mood, time spent, and post-usage views.

8 Conducting users studies of characters

Information technology is not what it used to be. The traditional function – a work-related tool with which to accomplish 'serious' tasks – still constitutes an important type of use, but it has been supplemented with a great number of other functions: education, service, entertainment, art, propaganda, social interaction and fun. Today, information technology works as sources of cognitive, social, emotional and aesthetic experiences in all kinds of context. The methods and goals of user studies have to adapt to the new conditions. 'Goodness' can no longer only be defined in terms of efficiency or ease of learning.

By introducing the notion of *experience*, we hope to have broadened the scope and aim of user studies. Experience involves complex conscious and non-conscious processes of cognition, emotions, and communication, of which both computer science and psychology have a rudimentary understanding. Furthermore, many of these processes are influenced by dispositions, expectations and everyday models of the world, which are determined not only by personal experience, but by the socio-cultural context. From our discussion above it becomes obvious that there is no way we can capture such a complex understanding of user reactions unless we bring into the study a larger picture of the contextual factors pertinent to the interpretation of the system. If we are to describe and understand these processes – as well as design for them – we have to integrate knowledge not only from cognitive psychology, but also from emotion theory, anthropology, communication studies, cinema studies and cultural studies. Although this broadening makes evaluations more intricate, it does not mean surrendering to the complexity. It only means that we have to take more of the context into account.

Furthermore, we also have to find innovative methods of measurements. In this paper, we have introduced a couple of light-weight parameters that we think should be worth developing and refining, for instance, the metaphor analysis combined with a careful reading of the inter-

views, and the statements that subjects could agree with.

In a way, we feel our approach presented here falls in between ethno-methodology and traditional usability studies. On the one hand, ethnographers thinks that human-computer interaction should spend less energy on putting figures and numbers on users and more on interpretative and qualitative analysis of interviews and actual behavior in the field. On the other hand, traditional usability studies might reproach us for spending time with a system that obviously does not promote wayfinding and task efficiency or can be measured in terms of those[6]. Still, user studies are needed in this, more complex, field as well. We need to provide input to the design of characters, and that input must not disregard important aspects of why the character succeeds or fails.

It is still early days for interface characters. Once they have been around for a while, there will be a better design basis for designers, but equally important is that users will have encountered them and started incorporate them into their view of what computer interfaces are and can be. Just as with any other mass-communication medium, producers of technology have to be aware of the fact that technology shapes people's view of the world, including technology. Machines with intentional processes (such as characters or other forms of intelligent user interfaces), that users can interact with, have never been possible before. Once they are, there will be room for ethnographic studies, but today we shall have to struggle with other kinds of studies like the one described here.

An important contribution from our study, is our attempt to broaden the view on what it means to create a life-like, believable, character. While animators in cartoon films are well aware of the need to create personality ('brain'), interface designers and researchers are still mostly concerned with how to generate facial expressions, body language, and coherent dialogues. An additional burden when designing interface characters compared to movies and novels is the *interactivity* between user and system. In a cartoon movie, the spectator is a passive viewer who has no influence over the course of actions. Our approach with introducing two characters that can talk between themselves in order to convey more of their attitudes and personalities to the users, is in our view an interesting way to combine a more passive spectator view with the interaction possibilities provided by computers. In our demonstration system, there were limited ways users could interact with the characters. Expanding on the interaction possibilities between user and character(s), while still keeping a strong narrative element and conveying the personality of the characters, is one of the interesting challenges to researchers in this field.

Acknowledgement

The work reported here was done as part of the EEC- and SITI-funded PERSONA project. Kristina Tullgren and Jenny Holm did the analysis of narrative and spatial verbs and adverbs as part of their Masters work. Jussi Karlgren provided some of the ideas for metaphor analysis. Anna-Lena Ereback helped us with the practical user study work. We thank them all, and in particular, we thank the anonymous participants in our study.

REFERENCES

André, Elisabeth, and Rist, Thomas (2000) Presenting through Performing: On the Use of Multiple Lifelike Characters in Knowledge-Based Presentation Systems, *Proceedings of Intelligent User Interfaces conference*, IUI'2000, New Orleans, Louisiana.

---

[6] Most traditional usability methods do include subjective comfort in addition to utility and efficiency, but their basis is in identifying users' task and making sure that they can be conduced efficiently.

Asimov, Isaac (1950) *I, Robot*, Gnome press, 1950. Doubleday, 1961, Bantam Books, 1994.

Benyon, D. and Höök, K. (1997). Navigation in Information spaces: supporting the individual, Paper presented at *INTERACT '97*, Australia.

Brennan, Susan E. and Ohaeri, Justina O. (1994) Effects of Message Style on Users' Attributions toward Agents, Conference Companion, CHI'94, Boston.

Bruner, Jerome (1986) Actual minds, possible worlds, Cambridge, MA, US, Harvard University Press.

Dahlbäck, N., Höök, K. and Sjölinder, M. (1996). Spatial Cognition in the Mind and in the World - the Case of Hypermedia Navigation, *The Eighteenth Annual Meeting of the Cognitive Science Society*, University of California, San Diego.

Dennett, Daniel Clement (1987) The intentional stance, Cambridge, Mass., MIT Press.

Elliott, Clark. (1992) The Affective Reasoner: A process model of emotions in a multi-agent system. Institute for the Learning Sciences Tech. Report #32.

Elliot, Clark and Brzeinski, Jacek (1998) Autonomous Agents as Synthetic Characters, Communications of the ACM, Summer 1998 issue.

Hayes-Roth, Gene Barbara, Ball, Christine Lisetti, Rosalind W. Picard, Andrew Stern (1998) Panel Session: Affect and Emotion in the User Interface at the Intellignet User Interface (IUI) conference, San Fransisco.

Höök, K. (2000) Steps to take before IUIs become real, Journal of *Interaction with Computers*, vol. 12, no. 4, February 2000.

King, W.J., and Ohya, J. (1995) *The representation of agents: a study of phenomena in virtual environments*, Proc. of the 4th IEEE International Workshop on Robot and Human Communication, RO-MAN'95, Tokyo, Japan.

Koda, T. and Maes, P. (1996) *Agents with Faces: The Effects of Personification of Agents*, Proceedings of HCI'96, London, U.K.

Lanier, J., (1996) My problems with agents, Wired.

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

Lakoff, G., and Johnson, M. (1999) *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*, Basic Books.

Lester, J., Converse, S., Stone, B., Kahler, S., and Barlow, T. (1997) Animated Pedagogical Agents and Problem-Solving Effectiveness: A Large-Scale Empirical Evaluation. In Proceedings of the *Eighth World Conference on Artificial Intelligence in Education*, Amsterdam, the Netherlands. IOS Press, pages 23-30.

Maes, P. (1994) Agents that Reduce Work and Information Overload, *Communications of the ACM*.

Maglio, P., and Matlock, T. (1999) "The Conceptual Structure of Information Space", In "Social Navigation of Information Space", A. Munro, K. Höök, and D. Benyon (eds.), Springer-Verlag, London.

Van Mulken, S., André, E., and Müller, J. (1998) The Persona Effect: How Substantial Is It. In Proceedings of HCI'98, pages 53-66, Sheffield, UK.

Murray, J. (1997) *Hamlet on the Holodeck: The Future of Narrative in Cyberspac*e, Cambridge: MIT Press.

Munro, A. J., Höök, K., and Benyon, D., eds. (1999) *Social Navigation of Information Space*, Springer September 1999 ISBN 1-85233-090-2.

Persson, P. (1999) Understanding Representations of Space: a Comparison of Visualisation Techniques in Mainstream Cinema and Computer Interfaces, in Munro, A., Höök, K. & Benyon, D. (eds.) *Social Navigation in Information Space*, London: Springer, pp. 195-216.

Paiva, Ana, and Martinho, Carlos (1999) A Cognitive Approach to Affective User Modeling, in the proceedings of the workshop on *Affect in Interactions; Towards a New Generation of Interfaces*, held during the i3 Annual Conference in Sienna, Italy.

Picard, Rosalind. (1997). *Affective Computing*. MIT Press.

Porter, Tom, and Susman, Galyn (2000) Creating Lifelike Characters in Pixar Movies, Communications of the ACM, January 2000, Vol. 43, No. 1, pp. 25-29.

Reeves, Byron and Nass, Clifford (1996) *The Media Equation: How People Treat Computers, Television, and New media Like Real People and Places*, Cambridge University Press.

J. Riseberg, J. Klein, R. Fernandez, and R. W. Picard (1997), Frustrating the User on Purpose: Using Biosignals in a Pilot Study to Detect the User's Emotional State, *CHI 98*, late-breaking short paper.

Shneiderman, Ben (1997) Direct Manipulation for Comprehensible, Predictable and Controllable User Interfaces, Proceedings of 1997 International Conference on Intelligent User Interfaces, (eds.) Johanna Moore, Ernest Edmonds, and Angel Puerta, ACM, Orlando, Florida.

Sproull, Lee, Subramani, M., Kiesler, S., Walker, J. H., and Waters, K. (1996) When the Interface is a Face, *Human-Computer Interaction*, 1996, Vol. 11, pp. 97 – 124.

Suchman, L. A. (1997) From Interactions to Integrations, *Proceedings of Human-Computer Interaction INTERACT'97*, S. Howard, J. Hammond, and G. Lindegaard (eds.), Chapman & Hall.

Tan, Ed S. (1996) *Emotion and the Structure of Narrative Film. Film as an Emotion Machine*, Mahwah: Lawrence Erlbaum Associates.

Waern, Annika, and Höök, Kristina (2000), Interface Agents: A new metaphor for human-computer interaction and its application to Universal Accessibility, in S. Constantie (ed.*) User interfaces for all*, Lawrence Erlbaum Associates.

Wright, Patricia, Belt, Steve, and Lickorish, Ann (1998) Animation, the fun factor and memory, proceedings of the *Computers and Fun* workshop in York, organised by Andrew Monk on behalf of the British HCI group.

Figures and tables

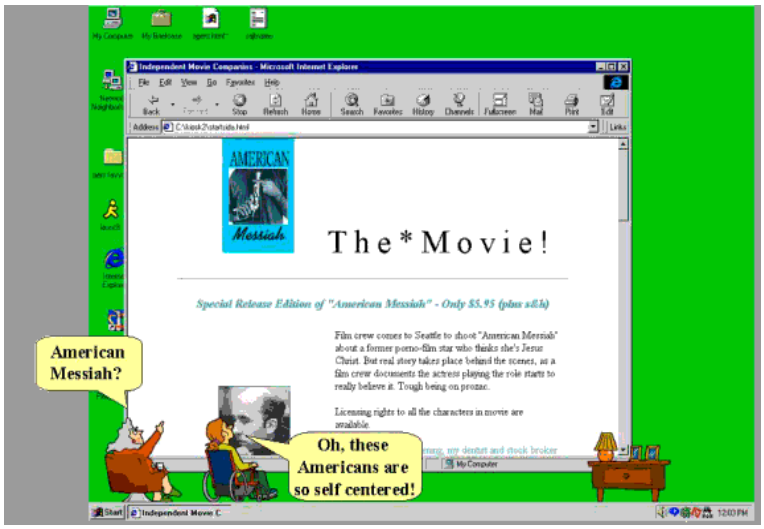

Figure 1. Agneta and Frida reacting to the site of a film production company.

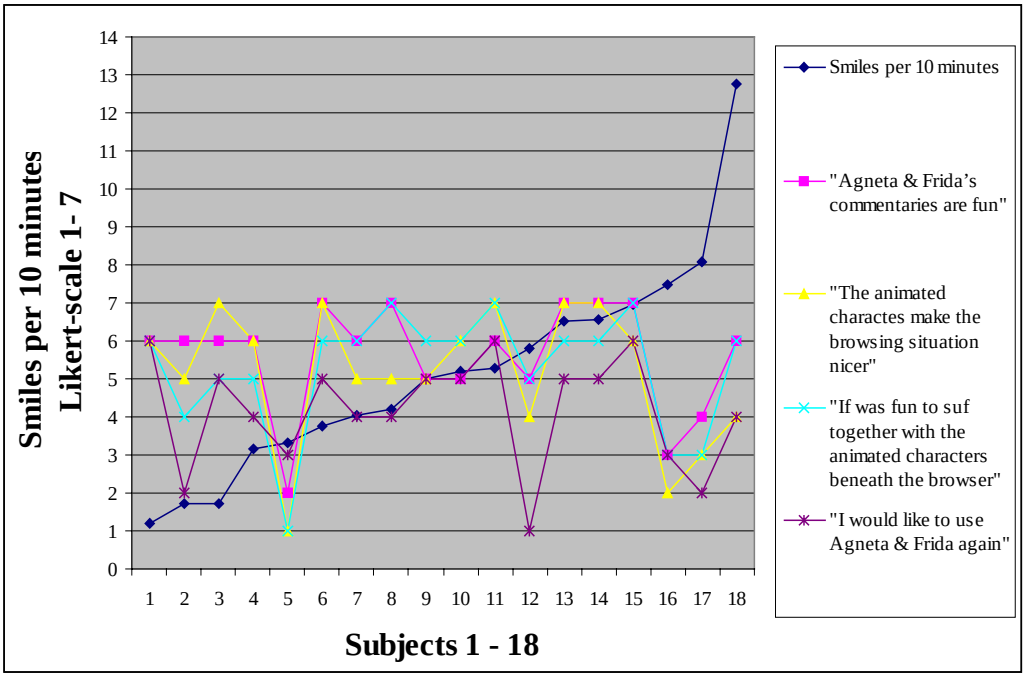

Table 1 Smiles/minute and the four 7-grade Likert-scale statements displayed in the same diagram for each subject, 1 - 18. The subjects are sorted by how many smiles they made from person 1 who only smiles 1.2 times per 10 minutes to subject 18 who smiled 13.3 times per 10 minutes.

Correlation Matrix

	again ok	nice ok	surf ok	fun ok
again ok	1.000	.699	.669	.468
nice ok	.699	1.000	.804	.815
surf ok	.669	.804	1.000	.790
fun ok	.468	.815	.790	1.000

Table 2 Correlation matrix for the questionnaire queries: “Agneta & Frida’s commentaries are fun” (fun), “The animated characters make the browsing situation nicer” (nice), “It was fun to surf together with the animated characters beneath the browser” (surf), and “I would like to use Agneta & Frida again” (again). $P < .05$.

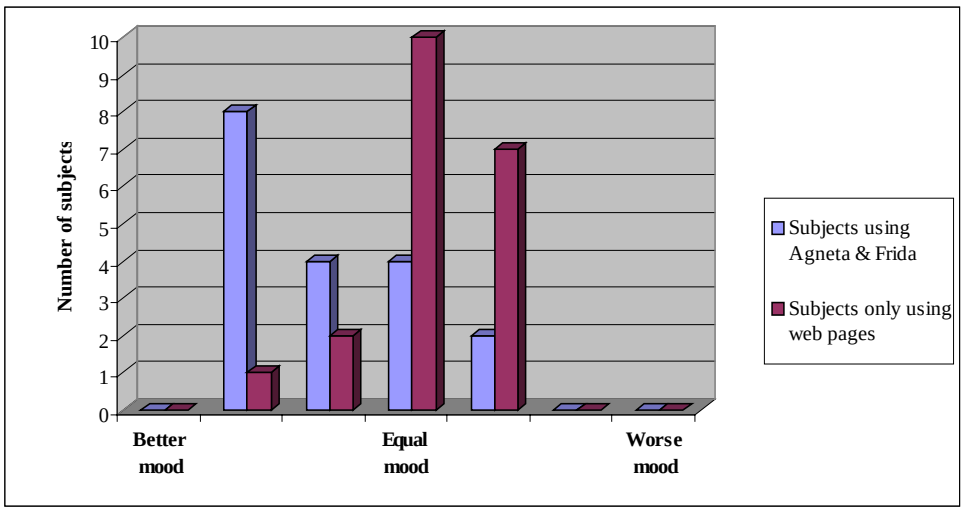

Table 3 Number of subjects who were in a better, equal or worse mood after using the Agneta & Frida system, versus only using the web pages.

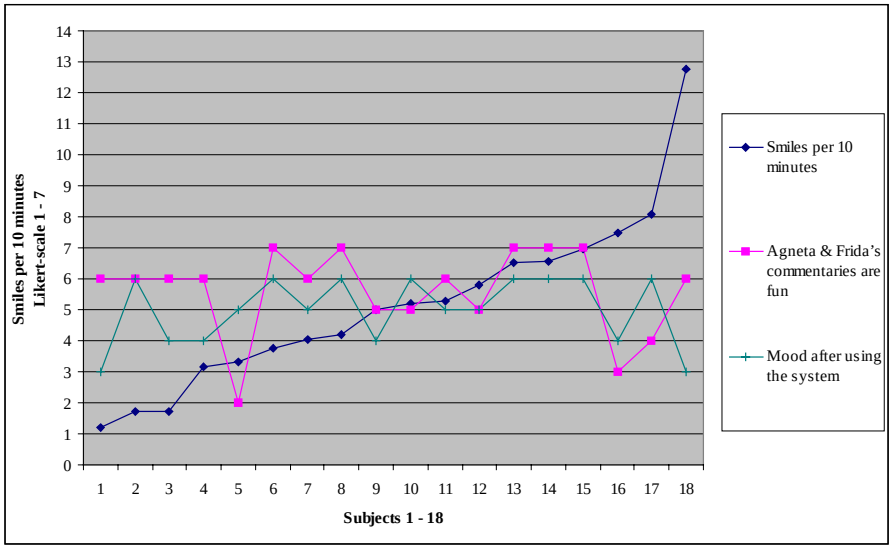

Table 4 Smiles per 10 minutes, one of the Likert-scale questions "Agneta & Frida's commentaries are fun", and subjects' mood after using the system scale 1 (very much worse) to 7 (much better mood) depicted in the same table. The 18 subjects are sorted by the number of smiles per 10 minutes. Again, none of the measurements follow one-another.

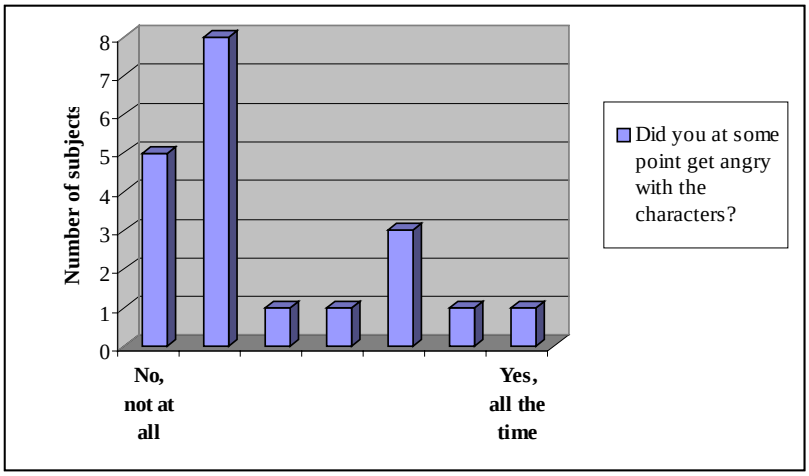

Table 5. Number of subjects who got angry with Agneta & Frida.

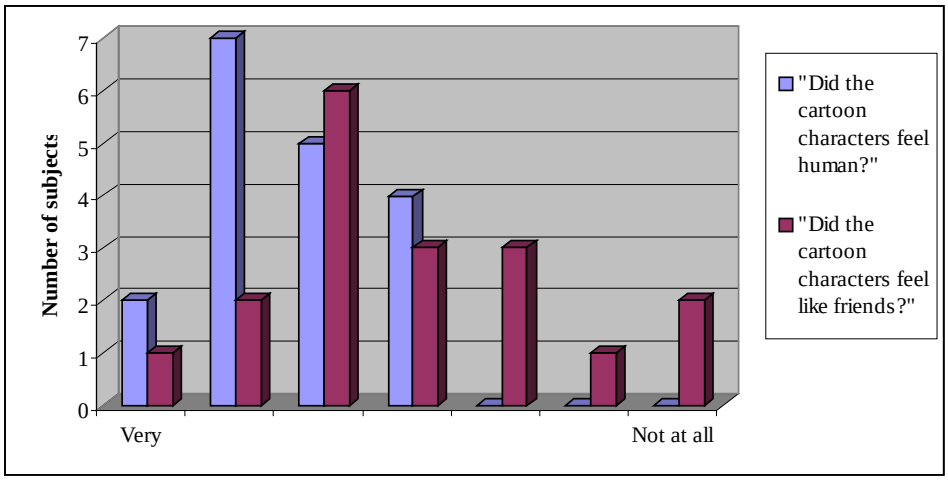

Table 6 Number of subjects who found Agneta & Frida to be human-like and number of subjects who found Agneta & Frida to be like friends..

	No of words in total	**No of spatial verbs + adverbs**	**No of narrative verbs + adverbs**	**No of narr**ative and spatial verbs + adverbs	Ratio of spatial versus narrative number of verbs + adverbs
Group A	4088	108	234	342	32% vs. 68%
Group B	6402	204	252	456	45% vs. 55%
Total	**10490**	**312**	**486**	**798**	

Table 7 Number of spatial versus narrative verbs and adverbs for group A and B.

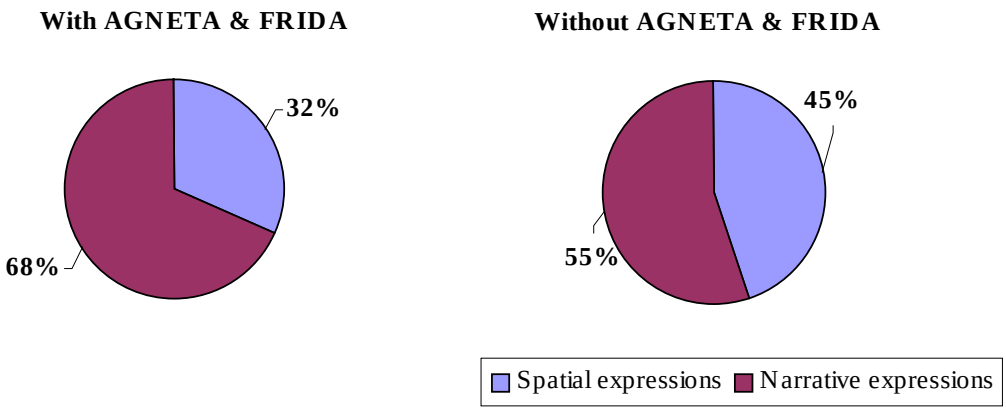

Table 8 Spatial versus narrative expressions with and without AGNETA & FRIDA.