



Realism is not all! User engagement with task-related interface characters

H.C. van Vugt^{a,b,*}, E.A. Konijn^a, J.F. Hoorn^b, I. Keur^a, A. Eliëns^b

^a Faculty of Social Sciences, Department of Communication Science, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

^b Faculty of Sciences, Department of Computer Science, Vrije Universiteit, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

Received 28 February 2006; received in revised form 10 July 2006; accepted 6 August 2006

Abstract

Human-like characters in the interface may evoke social responses in users, and literature suggests that realism is the most important factor herein. However, the effects of interface characters on the user are not well understood. We developed an integrative framework, called I-PEFiC, to explain ‘persona’ and realism effects on the user. We tested an important part of the model using an experimental design in which 140 middle school students were class-wise shown an informative virtual reality demonstration that incorporated either a realistic or an unrealistic (fantasy) interface character, or no character. Findings show, first, no persona effect on task performance. We discuss how user engagement might be related to persona effects. Second, designed realism of the interface character contributed to user engagement when controlled for various user perceptions. Moreover, perceived aesthetics and task-relevance further influenced user engagement. Third, user engagement and task performance combined better predicted satisfaction than either one of the factors alone. In sum, several appearance- and task-related factors contributed to user engagement and user satisfaction. Thus, realism is not all.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Human–computer interaction; Media entertainment; Empirical study; Interface characters; I-PEFiC model; Persona effect; Character realism; Engagement

1. Introduction

People can react to computers as they would react to real people (e.g., Nass et al., 1994, 1995; Reeves and Nass, 1996). An example of a user treating a human-like computer application in a social way, is: ‘You are very rude! You interrupt our conversation without any real reason. I’ll leave you, as you don’t wish to talk to me any more.’ (in De Rosis et al., 2005). Computer applications designed to resemble humans or animals have qualities, such as a human- or animal-like appearance, that are especially likely to elicit social responses. Of these computer applications, interface characters are a special type. By interface characters we refer to a broad range of characters that

may appear on the computer interface with a life-like visual appearance. They may be computer programs themselves and possibly behave intelligently (e.g., embodied conversational agents), or they may be part of a computer program (e.g., characters in games or desktop applications). They may or may not use gestures and speech to communicate and show emotions (e.g., Brave et al., 2005; Picard and Klein, 2002; Gratch and Marsella, 2005). Human–character interaction, then, is a type of human–computer interaction in which a human interacts (more or less actively) with an interface character.

Indeed, users respond differently to interfaces with an interface character than to those without an interface character, a phenomenon known as the ‘persona effect’ (Lester et al., 1997). However, empirical investigations often produce conflicting results concerning persona effects on the user (Dehn and Van Mulken, 2000). Several theoretical views address the persona effect on aspects of task

* Corresponding author. Tel.: +31 20 598 6839; fax: +31 20 598 6820.
E-mail address: hc.van.vugt@fsw.vu.nl (H.C. van Vugt).

performance and liking. In this introduction, we outline three of these theories and then propose a new theoretical approach to studying the topic in more depth.

The first theory states that interface characters may steer user attention, especially when they have an eye-catching appearance. This may improve task performance and liking. Some studies indeed found that an interface character increases a user's likeability of the system or enhances a user's learning experience (e.g., Moundridou and Virvou, 2002; Koda and Maes, 1996; Lester et al., 1997). Others found the reverse effect (e.g., Sproull et al., 1996); interface characters may distract users from their task, which may worsen task performance and liking (e.g., Takeuchi and Naito, 1995).

Second, social presence theory explains that in real life, the mere presence of another person increases anxiety and lowers levels of performance on complex tasks, but might facilitate simple task performance (Rickenberg and Reeves, 2000). Similarly, interface characters might evoke feelings of presence of 'another person', which might consequently affect task performance either positively or negatively, depending on task difficulty (e.g., Choi et al., 2001; Rickenberg and Reeves, 2000; Nowak and Biocca, 2003). In line with these ideas, Rickenberg and Reeves (2000) found that a 'monitoring' interface character led to decreased task performance in complex tasks. On the other hand, Beun et al. (2003) found that people had better memory recall when a character was present in the interface. In addition, some studies did not find a persona effect on task performance or recall at all. Van Mulken et al. (1998), for example, found that character presence had neither a positive nor a negative effect on comprehension and recall. Likewise, no persona effects were found in a tutoring domain (Moundridou and Virvou, 2002) and in an advice giving domain (Xiao et al., 2004).

Third, from a designer's point of view, the differentiating factor between character interfaces and traditional interfaces is human- or animal-likeness. Obviously, then, the degree to which the character resembles a real person or animal, in form and behavior (e.g., Bailenson et al., in press), is likely to influence the user. Form realism was subject of investigation in several studies, by comparing realistic and unrealistic outer appearances of interface characters. Koda and Maes (1996) found that realistic human faces are preferred over cartoon faces in terms of likeability and comfort. On the other hand, Catrambone et al. (2002) found that character realism (lifelike versus iconic) had little effect on the perception of an interface character, but user perceptions were strongly influenced by the task. Despite the contradictory findings, Dehn and Van Mulken (2000) conclude that realism is likely to have an important effect on users' responses. For this reason, the present study addresses realism.

Overviews of relevant factors (e.g., Ruttkay et al., 2004; Catrambone et al., 2004) indicate that factors related to the character's outer appearance, as well as factors related to the character's behavior, the user, and the task, all poten-

tially explain user responses. Therefore, this study takes a more fine-grained perspective by considering a variety of factors mentioned in the literature. In this paper, we present a model as a conceptual framework to explain liking of and user engagement with interface characters. Such a systematic integration of factors was missing in literature thus far. The model is user-centered, as it focuses on how users perceive interface characters within a task context. The opposite is a designer-centered perspective which may focus on the design of visual life-like appearances, communication modalities, behavior and emotion modeling.

User perceptions might depend on specific design characteristics (e.g., communication modality) and task context characteristics (e.g., the user's task goal). For example, an interface character using speech may be perceived as more realistic than an interface character which is text-based. Or, a funny interface character may be perceived as relevant when the user wants to be entertained, and irrelevant when the user wants to finish a task efficiently. We believe that this user-centered perspective provides us with a deeper understanding of the psychological processes underlying human-character interactions. Further, we believe the integrative model will be useful for the HCI community, and especially those dealing with interface characters, as it allows for a systematic empirical investigation of relevant factors, investigations into the relative importance of factors, and an integration of past research results. In addition, results can be used to inform the design of interface characters.

Interactive PEFiC

Social science theories can be applied to study human-computer interactions, because they resemble human-human interactions (cf. *the media equation*, Nass et al., 1994, 1995; Reeves and Nass, 1996). In line with these ideas, we take a new scientific perspective on studying humans interacting with interface characters. We apply a specific theory based on interpersonal attraction, psychological theories of emotion, and media entertainment, which explains user engagement with fictional characters, such as film and television characters (Konijn and Hoorn, 2005). This model on Perceiving and Experiencing Fictional Characters (PEFiC) (Hoorn and Konijn, 2003; Konijn and Hoorn, 2005) integrates several factors that contribute to human engagement with fictional characters. We may conceive of interface characters in a way similar to film characters.

Unlike film characters, however, people can actively interact with interface characters. To account for this *interactive* nature, the PEFiC model was adjusted to include engagement during human-character interaction, as well as interface character use. The model is called *Interactive PEFiC*, which is abbreviated to *I-PEFiC*, see Fig. 1 (Van Vugt et al., 2004, 2006a). In the present study, we focus on the engagement process evoked during human-character interaction, whereas the lower part of the model (affordances, use intentions, and use) is addressed elsewhere (Van Vugt et al., 2006a,b).

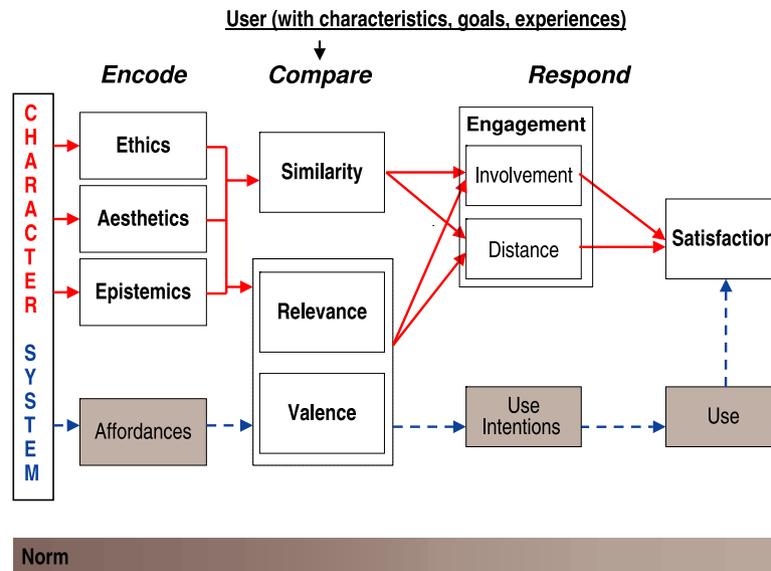


Fig. 1. Graphical representation of the I-PEFiC model. Note: The gray areas and dashed lines indicate factors and processes that are not investigated in the present study (see Van Vugt et al., 2006a).

In analyzing the user's experience towards an interface character, I-PEFiC distinguishes between three phases, *encode*, *compare* and *respond* (see Fig. 1). Typical factors in the encode phase of character engagement, each modeled with a positive and negative dimension, are Epistemics (realistic vs. unrealistic), Aesthetics (beautiful vs. ugly), and Ethics (good vs. bad). *Epistemics* relates to the extent to which the interface character is perceived as a realistic representation of a real life character. Both form and behavioral realism may contribute to perceived realism (Bailenson et al., in press; Guadagno et al., in press). Behavioral realism of interface characters is concerned with, for example, the character's facial expressions, body and head movements, gestures, eye contact and gaze (e.g., De Rosis et al., 2003; Bailenson et al., 2001; Cassell et al., 1994; Cassell and Thórisson, 1999), as well as the character's abilities, intelligence, conversational and social behavior (Hayes-Roth, 2003; Dehn and Van Mulken, 2000). Form realism is concerned with the outer appearance of the interface character. Does it resemble an existing living creature, whether human- or animal-like, or is it an unrealistic fantasy creature? Form realism is not likely to affect the dynamics of user interactions with an interface character as much as behavioral realism. Form realism can, on the other hand, be important in terms of social identity, and hence engagement with the character. Dryer (1999), for example, contends that visual appearance, in terms of form realism, influences whether an interface character is perceived as agreeable (i.e., involving) or disagreeable (i.e., distancing), which is in line with our I-PEFiC model. In our study, we are interested in how form realism affects engagement with the character, while controlling for other factors.

Ethics relates to how morally good or bad the interface character is. For example, does the interface character have

bad intentions when communicating (e.g., spam)? Does the interface character comment ironically or even sarcastically on the user's actions (cf. Agnetta and Frida, see Persson, 1999)? Or, does the character provide kind, helpful instructions (e.g., Clippy?)¹. In our study we kept this variable constant across conditions. We only used characters that were helpful, and hence ethically good.

Aesthetics concerns the attractiveness of the appearance of the interface character (is she a beauty or an ugly character?). There are universal standards of beauty (e.g., average face shape and symmetry; see, among others, Johnston and Oliver-Rodriguez, 1997), that apply to real people. Similar standards probably apply to interface characters. Ugliness is induced by deviations from the beauty standards, for instance, interface characters having misshapen skulls or showing signs of 'physical' decay. To avoid unintended influence from aesthetics in the present study, we attempted to keep this variable constant across conditions, and we checked it post hoc.

The comparison phase entails, first, establishing one's personal relevance and valence towards the character, which is probably the core of establishing user engagement with interface characters (cf. Konijn and Hoorn, 2005). These factors are intertwined with the task-context in which users find themselves during human-character interaction. Features of the interface character are evaluated for their relevance (importance) to task goals of the user, and for their potential to hinder or sustain one's task goals (valence). Examples of general goals in human-character interaction are entertainment, efficient task completion, or learning. In an interface character context, valence is concerned with questions such as 'If I use the character, I

¹ Clippy is a product of Microsoft Inc. (<http://www.microsoft.com>).

will have fun' versus 'If I use the character, I will waste time.' Furthermore, similarity between the interface character and the user also influences user engagement. Characters that are more similar to the user (e.g., in looks and behavior) are likely to be more involving than dissimilar characters. Bailenson et al. (2001) contends that users treat interface characters embodied with their virtual selves fundamentally different from interface characters embodied with virtual others. In sum, the comparison factors of relevance, valence, and similarity impinge upon user engagement with the interface character.

The response phase, finally, concerns engagement with a character that consists of parallel tendencies to approach and avoid the character. This is the backbone of the processes of *involvement* (e.g., empathy, sympathy, challenge) and *distance* (e.g., antipathy, irritation, boredom). Konijn and Hoorn (2005) and Konijn and Bushman (in press) provide evidence that liking a mediated person is best explained by *both* involvement *and* distance experiences. Thus, involvement and distance are two distinct experiences that are not the ends of a continuum; they can be experienced at the same time. A user might feel sympathy for a character and at the same time think the character is boring (for example, a virtual news-reader). Therefore, the I-PEFiC model predicts that *satisfaction* with the character is fed by the simultaneously active processes of getting involved with and keeping distant from the character.

Because the PEFiC-part of the model was originally based on non-interactive engagements with fictional characters, I-PEFiC does not take specific task contexts into account. However, interacting with interface characters is usually directed to accomplish a certain goal through tasks, for instance, via the Internet to book a flight. It is well-known that in a task context, satisfaction is related to task performance (e.g., Gelderman, 1998). Hence, user satisfaction may not only depend on engagement with the interface character, but also on task performance. Furthermore, engagement might also directly affect task performance. The assumption is that increased engagement with an interface character improves task performance when the character's actions are useful for the user's task. Increased engagement may worsen task performance when the interface character is distracting the user from the task. Therefore, the present study explores a variety of factors that contribute to engagement with a task-related interface character, satisfaction with that character, and task performance.

In general, we expect that positive dimensions of the factors (e.g., realism, beauty) result in positive effects (e.g., involvement) and negative dimensions (e.g., unrealism, ugliness) result in negative effects (e.g., distance). However, the model also accounts for the positive effects of negative dimensions as well as negative effects of positive dimensions. This may explain why interface characters can, for example, be ugly (negative dimension) and involving (positive effect) at the same time. For example, unexperienced users may like Clippy because it tries to be helpful, even if they find it ugly. Furthermore, Konijn and Hoorn

(2005) provide evidence that both realistic and unrealistic fictional characters evoke involvement, but realistic characters more so than unrealistic ones. Similarly, both realistic and unrealistic characters evoke distance, but unrealistic characters more than realistic characters. Therefore, the I-PEFiC model serves as a theoretical framework to explain potential persona and realism effects on engagement, satisfaction and task performance as studied in the present study. The model takes a user-centered perspective by focusing on user perceptions, rather than focusing on design characteristics.

2. Hypotheses

We designed an experiment to study the effects of 'designed realism' by taking a user-centered perspective. Using the I-PEFiC model as our theoretical framework, we focus on user perceptions, including perceived realism and perceived aesthetics, to understand how user engagement is evoked. Task performance and user satisfaction will be studied as effect variables. With this study we aim to increase the understanding of the contradictory findings found in the literature thus far.

Persona hypothesis

Although the literature is unclear about whether a 'persona effect' on task performance exists, we hypothesized affirmatively. In general, we expect positive effects on task performance in particular when the character provides task-relevant information.

H1 The presence of an interface character positively affects task performance.

Realism hypotheses

Based on previous research, Dehn and Van Mulken (2000) stated that character realism was likely to be an important factor influencing user experiences. Specified according to earlier findings within the context of the PEFiC model (Konijn and Hoorn, 2005; Konijn and Bushman, in press), we formulated two hypotheses:

H2a Both realistic and unrealistic characters evoke involvement, but realistic characters more so than unrealistic characters.

H2b Both realistic and unrealistic characters evoke distance, but unrealistic characters more so than realistic characters.

Satisfaction hypothesis

In addition, we are interested in relations between user engagement, task performance, and user *satisfaction*. The I-PEFiC model predicts that satisfaction is a function of involvement and distance, but in a task context, satisfaction is also highly related to task performance (Gelderman, 1998). Therefore, we formulated the following hypothesis:

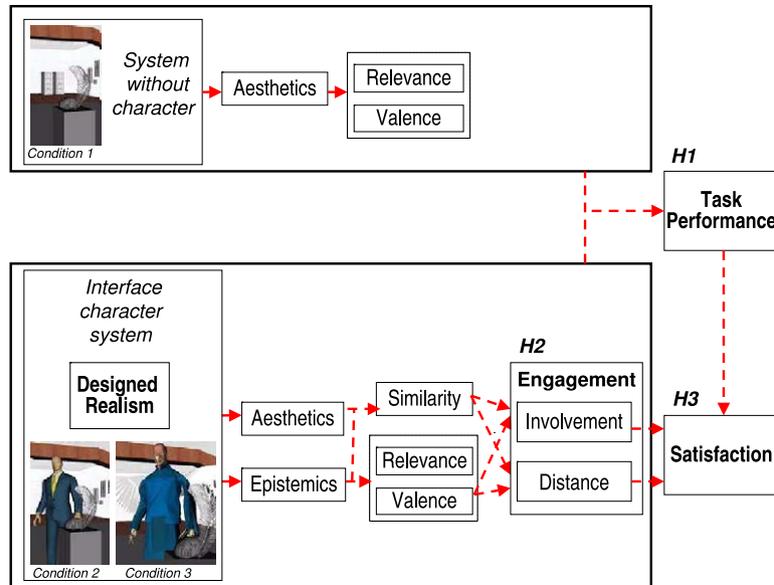


Fig. 2. Graphical representation of the three experimental conditions in the present study.

H3 Engagement and task performance both contribute to user satisfaction. In coalition, they better explain user satisfaction than either of the variables alone.

3. Method

3.1. Design

To address our core hypotheses adequately, we created a between-subjects design consisting of three conditions: (1) system without interface character (2) system with a realistic interface character (3) system with an unrealistic interface character. Fig. 2 is a graphical representation of the relatively complex experimental design. The conditions 1–3 and hypotheses H1 to H3 are indicated in the figure. Obviously, the variables epistemics (realism), similarity, and engagement (involvement and distance) cannot be studied in condition 1 (the system without an interface character). As can be seen in Fig. 2, the dependent variables task performance and satisfaction were measured in all conditions. Furthermore, the control variables perceived aesthetics, perceived (task-) relevance, and perceived valence were included in *all* conditions.

3.2. Participants

The participants were adolescents (aged 12–18) in six classes at a Dutch middle school ($N = 140$). The number of participants varied from 16 to 32 participants per class, with a mean of 23. All participants used the computer at least weekly. Most participants (72%) played games on computers, boys (89%) more than girls (56%). The number of participants and the descriptive statistics on age and gender per condition can be found in Table 1.

Table 1

Participants' characteristics per experimental condition

| | (1) System without interface character | (2) Realistic interface character | (3) Unrealistic interface character |
|------------------------|--|-----------------------------------|-------------------------------------|
| Number of participants | 38 | 57 | 45 |
| M (SD) of age | 16.33 (.68) | 14.21 (1.69) | 13.61 (.62) |
| Percentage boys (%) | 32 | 47 | 56 |

3.3. Materials

Character features

To avoid problems of incomparability, one should keep the various experimental conditions as similar as possible. Therefore, we designed the interface characters ourselves, only varying the factor under study, that is, realism. We manipulated realism along the appearance dimension: the only differences between the unrealistic and realistic character were that the unrealistic character had a beak, wings and claws, which made it a fantasy creature (see Fig. 3). The posture, shape, color of the clothing, and movements of the two interface characters were equal.

Virtual environment

The stimulus materials were based on a 3D virtual environment deployed in a project performed in cooperation with the Dutch Cultural Heritage Institute (ICN)² in the context of the International Network for the Conservation of Contemporary Art (INCCA).³ In short, the idea was to

² <http://www.icn.nl/>.

³ <http://www.incca.org/>.

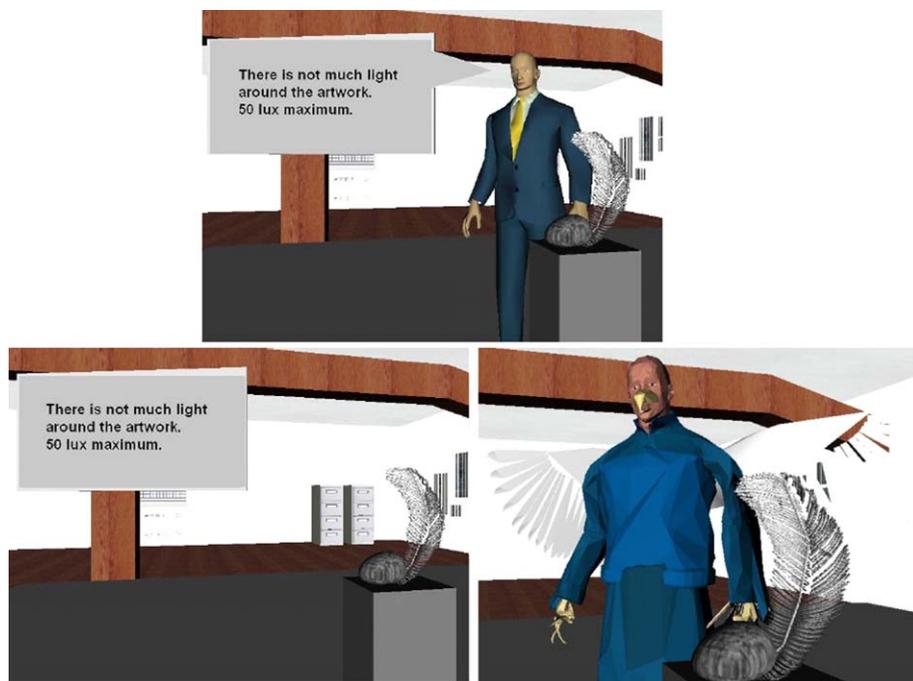


Fig. 3. The stimulus materials. The realistic character, with text balloon, next to ‘Stone and Feather’ (on top). The unrealistic character, with wings, beak and claw (right). The no-character version, with text frame only (left). The text is translated from the Dutch.

develop ‘digital dossiers’ (Hoorn et al., 2004) for individual artworks, allowing professionals to deal with the information in an integrated, interactive fashion. We developed a prototype of a VR (the artist’s studio) that contained information on the artwork ‘Stone and Feather’ of the Dutch artist Marinus Boezem. The environment was multimedia-enhanced; it included visual (e.g., the artwork, photographs), textual (background information), and audio (e.g., audio–video recording) information. The environment allowed for the inclusion of the artist in the form of an interface character. The VRML based DLP + X3D platform was used to create the environment, and the DLP + X3D/STEP agent platform to create interface characters that move naturally and use complex gestures (Eliëns et al., 2002). For more technical details, we refer to Huang et al. (2003).

Demonstration materials

For the experiment, we developed three demonstration versions (no-character, realistic character, unrealistic character) of the 3D virtual environment viewable with Windows Media Player. The demonstration showed the construction of an artwork. The viewer was led through the environment while text pointed out various details of construction and the characteristics of the artwork (e.g., ‘The feather is an ostrich feather’). Thus, the same communication modality (text) was used in the character conditions as in the no-character version to avoid that communication modality (e.g., speech versus text) might explain the effects (e.g., in studies of Sproull et al., 1996; Moundridou and Virvou, 2002). This textual information was given either as a text balloon in the embodiment conditions, or as a text frame in the no-character condition.

These balloons and frames had equal sizes (see Fig. 3). Each version included nine text balloons.

The size of the text and the presentation speed made the demonstration easily readable from several meters away. No sound accompanied the demonstration to avoid possible confounding. The total duration of the demonstration was about 4 minutes. We used a laptop (ASUS 5600N), a projector, and a big screen (about 150 by 200 cm) to show the demonstration to the participants.

3.4. Procedure

At a middle school, six classes cooperated in the research that was presented to the students as ‘aimed at improving computer-mediated education’. The classes participated during one day, sequentially, all in the same classroom. They were instructed to watch the demonstration silently. To ensure that students paid attention to the demonstration, we told them that we would return the next day to see how well they could reconstruct the piece of art. However, as our dependent measure, the participants completed a recall test that was related to the actual reconstruction task. Directly after watching the demonstration, the participants completed this recall test as well as a user perception questionnaire (see Section 3.5). After they finished, the participants received a small present. The next day, the participants were debriefed and thanked again for participating in the experiment. The teachers explained that the instructions concerning the reconstruction of the artwork were mainly used to guide their attention, and that it would not be necessary to actually reconstruct the artwork. Students were also told that they could contact the researchers

anytime for more information concerning the artwork and the experiment.

3.5. Measurements

All measurements were taken by means of paper-and-pencil questionnaires. We developed a recall test to measure task performance. A structured questionnaire was used to measure user perceptions, engagement, and satisfaction.

The *recall test* consisted of 10 multiple-choice questions about the information given in the demonstration. Task performance was assessed as a percentage of correct responses. Prior to the experiment, we performed a pilot test to assess task and test difficulty. A variation in correct answers told us that the test was neither too difficult nor too simple, and that the language and questions used were comprehensible for the target test population.

The structured *user perception questionnaire* contained Likert-type scales. Each item was followed by a 6 point rating scale, ranging from 1 (do not agree at all), 2 (do not agree), 3 (barely agree), 4 (agree a little), 5 (agree), to 6 (fully agree). Items were presented in mixed order and existing scales were used wherever possible. When necessary, items were translated, adjusted to the purpose of the study, the specific material (interface characters), and the language use of the target group of adolescents. In order to avoid directing participants in an affirmative direction (see Dillman, 2000), all scales contained indicative items (e.g., 'The figure looks natural') and counter-indicative items (e.g., 'The figure looks fake').

As an indication of how well a set of items measures a latent construct, we use Cronbach's α (Cronbach's α). In other words, Cronbach's α indicates whether a set of items is a homogeneous set that covers the meaning of the theoretical construct. The higher the Cronbach's α , the more reliable the generated scale is. A Cronbach's α above .60 generally indicates an acceptable scale (Dillman, 2000).

Perceived realism was measured with 4 items (cf. Konijn and Hoorn, 2005). A Cronbach's α of .55 indicated that the statistical reliability of the perceived realism scale was marginal. Psychometric analyses gave no indicators to improve this scale. Although group-wise comparisons are still allowed, results should be interpreted with caution.

The measurement of *engagement* was based on scales used by Konijn and Hoorn (2005), distinguishing between *involvement* and *distance*. The involvement scale was measured using two items ('The figure is appealing' and 'The figure makes me happy') and was reliable according to a Cronbach's α of .64. The distance scale was measured using two items ('The figure is annoying' and 'The figure irritates me') and was reliable according to a Cronbach's α of .77.

The scales for *perceived aesthetics*, *perceived relevance*, and *perceived similarity* were also based on Konijn and Hoorn (2005). The perceived aesthetics scale consisted of two indicative items (e.g., 'The figure is beautiful') and two counter-indicative items (e.g., 'The figure is ugly').

Compared to the original scale, one item had to be discarded because of poor fit (i.e., an inter-item total correlation $<.30$). The remaining scale was reliable (Cronbach's $\alpha = .77$). The perceived relevance scale consisted of two indicative items (e.g., 'What the figure says, is important') and two counter-indicative items (e.g., 'If something is said in the demo, I think: never mind!'), and had a satisfactory reliability (Cronbach's $\alpha = .79$). The perceived similarity scale comprised of only two items, one indicative item (e.g., 'I resemble the figure') and one counter-indicative item (e.g., 'The figure and I are different'). This scale was reliable (Cronbach's $\alpha = .82$).

In addition, *perceived valence* and *satisfaction* items were based on technology acceptance theory (e.g., Davis, 1989; Venkatesh et al., 2003) and requirements analysis literature (e.g., Hoorn et al., in press). Perceived valence was measured using eight items. We included various items to test whether different types of valence, concerning effectiveness, accuracy, speed and effort, would all fit into one valence-scale. The perceived valence scale consisted of four indicative items (e.g., 'I can reproduce the artwork without errors', i.e., accuracy valence) and four counter-indicative items (e.g., 'It is difficult to reproduce the artwork', i.e., effort valence). The different types of valence fit into one scale that was reliable according to a Cronbach's α of .90. The satisfaction scale consisted of two indicative items (e.g., 'The figure was clear') and two counter-indicative items (e.g., 'The figure was confusing'). A Cronbach's α of .80 indicated a reliable scale to measure satisfaction.

The scales that dealt with specific character properties (realism, similarity, and engagement) could not be included in the questionnaire used in the no-character condition. Where needed, items containing 'figure' were replaced by items containing 'demo'. For example, 'The demo is beautiful' (perceived aesthetics) and 'What is said in the demo, is important' (perceived relevance). The questionnaire used in the character conditions consisted of 30 items, and the questionnaire in the no-character condition consisted of 20 items. Additional information was collected about the students' gender, age, game usage (hours per week), computer usage (hours per week), and school level (either low or high).

Appendix A shows the recall test, translated from Dutch. Appendix B shows the items of the user perception questionnaires and the reliabilities of each scale.

4. Results

In this section, we first check whether the realistic character was indeed perceived as more realistic than the unrealistic character (manipulation check). After that, we test our hypotheses and perform additional analyses to look into user perceptions. The criterion for significance was set as $\alpha = .05$. Note that the persona hypothesis is tested on task performance as dependent variable and the realism hypothesis on engagement (involvement and distance) as

Table 2
Means and standard deviations (SD) of perceived realism, aesthetics, relevance, valence, similarity, involvement, distance and satisfaction per experimental condition

| | Perceived realism | Perceived aesthetics | Perceived relevance | Perceived valence | Perceived similarity | Involvement | Distance | Satisfaction |
|-----------------------|-------------------|----------------------|---------------------|-------------------|----------------------|-------------|-------------|--------------|
| No character | n.a. | 3.12 (1.17) | 2.96 (1.13) | 3.52 (1.26) | n.a. | n.a. | n.a. | 3.51 (1.46) |
| Realistic character | 3.09 (.98) | 2.87 (1.07) | 3.35 (1.18) | 3.22 (1.07) | 1.81 (1.06) | 2.11 (.99) | 3.66 (1.38) | 3.81 (1.03) |
| Unrealistic character | 2.62 (.90) | 2.65 (1.07) | 3.87 (1.19) | 3.47 (.94) | 1.53 (.82) | 2.41 (.98) | 3.58 (1.23) | 3.64 (.96) |

n.a., non applicable.

dependent variable. Finally, these dependents are tested for their combined contribution to satisfaction.

4.1. Manipulation check

To assess the success of our realism manipulations, an analysis of variance (ANOVA) with designed realism (realistic versus unrealistic) as between subjects variable and perceived realism as dependent variable was performed. Participants perceived the realistic character as significantly more realistic ($M = 3.09$; $SD = .98$) than the unrealistic character ($M = 2.62$; $SD = .80$): $F(1,100) = 5.90$, $p < .02$. The successful manipulation justified further hypotheses testing.

4.2. Testing hypotheses

To test whether the presence of an interface character affected task performance (here, recall) or not (H1, persona hypothesis), we conducted a contrast analysis of variance (ANOVA). This is a solid method to compare the mean of one level (no-character) to the mean of other levels combined (realistic and unrealistic character). Results showed that the no-character ($M = 81\%$ correct answers; $SD = 17$) and the interface character ($M = 78\%$ correct answers; $SD = 17$) conditions led to similar levels of recall ($p = .63$). Thus, we did not find support for the persona hypothesis (H1) as we did not find a persona effect on task performance.

Second, to test whether the realism hypotheses (H2a and H2b) were supported by the empirical data, a multivariate analysis of variance (MANOVA) was conducted. We entered designed realism (realistic vs. unrealistic) as the between-subjects factor, and involvement and distance as the dependent variables. Results showed that the realistic and unrealistic character evoked similar levels ($p = .31$) of involvement ($M(\text{realistic}) = 2.11$; $SD = .99$, and $M(\text{unrealistic}) = 2.41$; $SD = .98$) and distance ($M(\text{realistic}) = 3.66$; $SD = 1.38$, and $M(\text{unrealistic}) = 3.58$; $SD = 1.23$). Thus, we did not find empirical evidence that supported these two hypotheses. Realistic characters did not evoke more involvement than unrealistic characters (H2a), and unrealistic characters did not evoke more distance than realistic characters (H2b).

Third, we tested whether the empirical data supported the satisfaction hypothesis (H3). We expected that satisfac-

tion using an interface character would be significantly influenced by task performance, involvement with, and distance toward the interface character. Therefore, we performed a regression analysis (method Enter), with task performance, involvement, and distance as predictors, and user satisfaction as the dependent variable. The model explained 26% of the variance in satisfaction. Distance appeared to be the best predictor of satisfaction with an interface character ($\beta = -.41$, $p < .001$, partial $r = -.40$, semi-partial $r = -.37$) followed by task performance ($\beta = .32$, $p < .01$ partial $r = .34$, semi-partial $r = .32$). Involvement did not significantly contribute to satisfaction ($\beta = .07$, $p = .42$). Based on these results, the satisfaction hypothesis can be partly sustained; in coalition, distance (as part of the engagement process) and task performance can better explain satisfaction than either of the variables alone.

4.3. Additional analyses

Thus far, we tested the effects of *designed* characteristics on users' responses. Yet, because the I-PEFiC model predicts that a number of user *perceptions* in the encoding stage are indicative to user engagement, we tested specific user perceptions in predicting user engagement. In addition, we tested the effects of user perceptions on task performance. Thus, we tested the persona and realism hypotheses again, but this time we included the user perceptions in the analyses, in addition to the persona factor and the realism factor. The participants' means on the various user perceptions, as well as on involvement and distance, are presented in Table 2.

First, we tested whether user perceptions could explain involvement and distance better than designed realism. Therefore, we performed an ANOVA with designed realism (realistic versus unrealistic) as the between-subjects factor, and involvement and distance as the dependents. Furthermore, we included perceived realism, perceived aesthetics, perceived similarity, perceived (task-)relevance, and perceived valence as covariates. The multivariate effects were tested using the multivariate criterion of Wilks' lambda. This statistic evaluates the multivariate hypothesis that the population means on the multiple dependent variables are equal across groups, while controlling for interdependency between dependents (e.g., Stevens, 2002). The multivariate test showed that designed realism significantly

affected the dependents, that is, involvement and distance taken together (Wilks' $\lambda = .93$, $F(2, 93) = 3.47$, $p < .04$, partial $\eta^2 = .07$).⁴ Furthermore, perceived aesthetics (Wilks' $\lambda = .67$, $F(2, 93) = 32.07$, $p < .001$, partial $\eta^2 = .33$) and perceived relevance (Wilks' $\lambda = .79$, $F(2, 93) = 12.28$, $p < .001$, partial $\eta^2 = .21$) also affected the dependents. Perceived similarity and perceived valence did not significantly affect the dependents ($ps > .05$).

More specifically, further tests specified whether designed realism and the user perceptions contributed to involvement or distance, or both. First, designed realism affected involvement ($F(1, 94) = 6.44$, $p < .02$, partial $\eta^2 = .06$). Of the user perceptions, only perceived aesthetics affected involvement ($F(1, 94) = 19.22$, $p < .001$, partial $\eta^2 = .17$). The more beautiful the character was perceived to be, the more involved participants were. Second, both perceived aesthetics ($F(1, 94) = 25.43$, $p < .001$, partial $\eta^2 = .21$) and perceived (task-)relevance ($F(1, 94) = 22.82$, $p < .001$, partial $\eta^2 = .20$) affected distance. The more beautiful and relevant the character was perceived to be, the less distant participants felt. Note that we did not find effects of *perceived* realism on either involvement or distance. However, designed realism did affect involvement. Thus, if we control for the user perceptions in the analyses, we found that realism affected user involvement which supports realism hypothesis H2a, in contrast to results in Section 4.2. Note that designed realism and perceived realism are highly related (we checked the manipulation of designed realism using the perceived realism scale), which makes it difficult to separate these effects. In sum, the results of this additional analysis showed that, next to designed realism, several user perceptions significantly contributed to user engagement. That is, perceived aesthetics affected both user involvement and distance towards the character, whereas perceived relevance contributed to user distance.

Second, we tested to what extent user perceptions could explain task performance. We performed a contrast analysis within ANOVA, comparing the no-character system with the interface character systems (character presence was the between subject factor), and task performance as the dependent variable. Furthermore, we entered perceived aesthetics, perceived relevance, and perceived valence as covariates. As in the foregoing analysis (concerning H1), we found that character presence did not affect task performance ($p = .46$). In addition, none of the user perceptions significantly influenced task performance ($ps > .05$), indicating that user perceptions (of interface characters to which users feel distant) were not decisive for task performance.

Then, we questioned whether task performance, involvement, and distance were affected by the personal characteristics of participants (gender, age, game usage, computer

usage, and school level). Therefore, we performed two additional analyses. An additional ANOVA was performed with personal characteristics as covariates to see whether these could explain task performance. We found a significant main effect of school level ($F(1, 121) = 12.61$, $p < .001$, partial $\eta^2 = .09$) on task performance. Higher qualified students performed better ($M = 82\%$ correct answers; $SD = 18$) than lower qualified students ($M = 74\%$ correct answers; $SD = 14$). This result follows general expectations and did not affect our theoretical assumptions.

An additional MANOVA was performed to test for effects of the personal characteristics on user involvement and distance. The multivariate test showed that only age affected (one of) the dependents (Wilks' $\lambda = .86$, $F(2, 88) = 7.30$, $p < .001$, partial $\eta^2 = .14$). Further tests showed that age affected distance ($F(1, 89) = 14.76$, $p < .001$, partial $\eta^2 = .14$), but not involvement ($p = .26$). The elder the students were, the more distant they felt towards the interface characters. Perhaps, the elder students might have found the demonstration childish or have a more negative attitude in general.

5. Conclusion and discussion

The main goal of the study was to investigate persona and realism effects on character engagement and task performance, taking a user-centered perspective. Hereby, we focused on form realism, not behavioral realism. In our study we tested an important part of the I-PEFiC model, which aims at explaining user engagement and satisfaction with interface characters. This model integrates a variety of factors mentioned in the literature, and focuses on how users perceive interface characters within a task context.

First, we did not find a persona effect on task performance. The presence of an interface character neither increased nor decreased task performance, as compared to the no-character system. In addition, user perceptions (of character's aesthetics, task relevance, etc.) did not contribute to task performance. Thus, task performance remained unaffected by the presence or absence of an interface character. Second, the results of hypothesis testing indicated no support for the realism hypothesis. That is, whether the interface character had a realistic or unrealistic design did not affect users' involvement nor distance with the character. However, if users' perceptions of the interface character were included in the analyses, designed realism did affect involvement (but not distance). This supports the realism hypothesis that (designed) realism of an interface character affects the users' involvement with that interface character. We did not find an effect of perceived realism, which might be due to the relatively low reliability of this scale.

Results further indicate that several user perceptions affected involvement and distance with interface characters. It appeared that users felt more involved with a character

⁴ Using partial eta-squared (partial η^2) as the measure of effect size. In general, the conventional cutoffs are .01, .06, and .14 for small, medium, and large η^2 , which are even too large for partial η^2 (Green and Salkind, 2005 p. 187).

when they perceived it as more beautiful. In addition, users felt less distant to a character when they perceived it as more beautiful and more relevant for the task. This is in line with results of Van Vugt et al. (2006a) in which engagement was affected by both the aesthetic looks of an interface character, as well as the relevance of the affordances the interface character offered for task completion. Last, we found support for the satisfaction hypothesis. Both task performance and distance predicted user satisfaction with an interface character. Involvement, however, did not significantly contribute to satisfaction. Participants felt rather distant to, and not so much involved with, both the realistic and the unrealistic interface character. This might explain why distance, and not involvement, was a predictor of satisfaction. This finding shows resemblances with results from Konijn and Hoorn (2005) and Konijn and Bushman (in press). They found that distance was a better predictor of liking a character than involvement. The result is also in line with Lindgaard and Dudek (2003) and Van Vugt et al. (2006a) who found that satisfaction is a subjective sum of several user experiences with the system. Both the appearance, as well as task-related features of an interface character affect user satisfaction with that character. Because end-user satisfaction is seen as an important goal in user-system interaction and design, our results show that it is important to enhance both engagement and features related to the task at hand.

The present results showed that perceived aesthetics was the most important variable that influenced user engagement with the interface character. The more beautiful users found the character, the more engaged they were. This is in line with results of Van Vugt et al. (2006a), in which aesthetics was also the best predictor of user engagement. The result is also in line with results from studies dealing with interpersonal attraction in real life, showing that beautiful people are generally preferred over ugly people. Furthermore, aesthetics also influences users when they are interacting with computers (e.g., Tractinsky et al., 2000). Perceived aesthetics might account for some unexpected results of previous studies on interface characters. For example, the study of Beun et al. (2003) experimentally compared three interfaces incorporating either a realistically designed, pretty female interface character, a cartoon-like character, or no character. Participants performed better on a memory task when the female character was present than when the cartoon-like or no character was present. Although these results were said to support the persona effect hypothesis on task (here, memory) performance, this was only true for the female interface character.

Using the insights of the present study, we may understand why interaction with the female character, but not with the cartoon-like character, led to higher recall than in the no-character condition. This might have been due to her aesthetic appeal. As we learned in the present study, perceived aesthetics of an interface character positively contributes to user engagement. It is likely that the female

character was more attractive to the (64% male) participants than the cartoon character, and hence more engaging (involving). Possibly, the engaged students paid more attention to the story told by the pretty interface lady than the students that used the cartoon-like or the no-character version. As a result they had a better recall. Thus, the positive effect of the character on task performance might have been the result of the particular character used, that is, a beautiful one.

Furthermore, in the study of Koda and Maes (1996) human and cartoon faces were used to study (form) realism effects. They stated that the level of realism of an interface character (referred to as level of abstraction or humanity) changes its likeability and perceived intelligence. However, it is unclear whether perceived realism was the factor that led to effects, or whether, in line with our study, perceived aesthetics was a decisive factor. After all, the human face could have been more aesthetically pleasing than the cartoon face. This is a rather fair explanation of the reported perceived intelligence effect, as empirical studies demonstrate that individuals perceive beautiful people to be more intelligent than ugly people (e.g., Kanazawa and Kovarb, 2004).

Remarkably, interface character research thus far has belittled the importance of aesthetics. This was also the case for a long time in other HCI fields (see Tractinsky, 1997; Liu, 2003; Norman, 2004). We believe that interface character research would benefit from a recognition of not only the importance of realism, but also of the importance of aesthetics in human-character interactions (see also Van Vugt et al., 2006a). How perceptions of realism and aesthetics of interface characters interrelate, and which factors contribute most to user engagement and satisfaction, should be investigated in future studies.

In the present study, not only perceived aesthetics but also perceived relevance of the character to accomplish a task was a predictor of engaging with the character. That is, when the interface character was perceived as more relevant to the task, the user felt less distant to the character. Unlike realism and aesthetics, relevance is a variable that is not clearly related to the character's appearance. However, this factor is related to the user's task goal. Is the character relevant for the user to perform the task in the particular environment? The importance of taking into account the task context in interface character research has been recognized (e.g., in Dehn and Van Mulken, 2000; Catrambone et al., 2002). Based on the results of our experiment within the I-PEFiC framework, we assume that it is not so much the task context itself, but the *perceived* relevance of the interface character in view of the task to be accomplished, that determines the users' responses. In a game, for example, the presence of an interface character that behaves realistically might be important to make the game entertaining (the user's goal). However, in an editing task, for example, such an interface character might be irrelevant for or even

obstructing efficient task completion (the user's goal). When taking into account the (perceived) relevance of interface characters in various task contexts, we might better understand why users react differently to the same character in different task contexts.

Given that various perception variables influence user engagement with interface characters, we might comprehend why some studies found a persona effect and other studies did not. Each study reported in the introduction of the present paper used different characters with different appearances and characteristics. In addition, the studies used different tasks of varying complexity. Thus, in the various studies, users might have perceived the interface characters differently in terms of aesthetics, realism, and relevance. Whereas a beautiful interface character providing relevant information for the user's task at hand may evoke strong user engagement, an ugly interface character not helping the user to perform his task may evoke no engagement at all. As a result, users might have been more engaged with a character in one study than in another study. Engaged users are likely to pay more attention to the character and experience more social presence (e.g., Choi et al., 2001). Thus, task performance and learning may increase, and persona effects can be found. Engaging interface characters might be advantageous in, for example, computer-mediated learning environments. In contrast, characters with which the user does not feel engaged, do not boost the attention or feelings of social presence of users, and hence, no persona effects are involved (such as in the present study). To study whether engagement is indeed a key factor in causing a persona effect, various levels of engagement with an interface character can be created. Then, we expect a persona effect when engagement is high, and no persona effect when engagement is low.

Ideally, stimuli differ only on *the one or a few* dimensions that are studied in a particular experimental setting, in order to attribute effects to that dimension(s). But, even when stimuli are designed to differ only in terms of realism, one cannot prevent other factors from playing a role as well. The reason is that users perceive characters on more than one dimension. These perceptions might suppress effects of the factor under investigation, or account for the effects themselves. Thus, it is important to control for factors that are likely to covary with factors under study, as we did in the present project. Overviews of relevant factors as well as our integrative I-PEFiC model can further inform future research.

Another point of methodological consideration is that our findings address a particular group of adolescents. We used a limited sample comprised of students of only one school and of a specific age range. Adolescents might care more about outer appearance in terms of aesthetics than adults do. This might partly explain the effects of aesthetics we found in the study. The participants also had a certain amount of computer experience and are used to playing games regularly. This might have biased their perceptions of the interface character. Often, students are

every-day users of top-of-the-art games featuring life-like characters in graphically rich environments having a high degree of behavioral realism. They might have found the character in the demonstration boring and dull, which would explain the high level of distance they felt towards the character. On the other hand, users with limited computer and game experience might find the character such as presented in the demonstration interesting.

Perceptions of realism are likely to be affected by both form and behavioral realism. We may question whether perceived realism will have a larger contribution to user engagement when behavioral realism is manipulated. In addition, we may question whether perceptions of aesthetics and task-relevance continue to be the main predictors of user engagement in other contexts. Further, we may question the generalizability of the results to real-life situations in which people actively interact with the interface character (e.g., Clippy or game characters). Then, users will experience the usability of the character software which concerns, for example, conversational capabilities of the character. This might not only increase the behavioral realism of the character, but also alter the outcome expectations (valence) as formulated in the I-PEFiC model. Also, the proactive and reactive behavior of a character might influence user responses in human–character interactions (cf. Xiao et al., 2003), as such a behavior might directly contribute to perceptions of realism, ethics, affordances, and valence.

In the present study, the I-PEFiC model provided a useful theoretical framework to determine factors that explain user responses to interface characters. Meanwhile, the I-PEFiC model has been tested using various interface characters in different task contexts. For example, we focused on situations in which users actively interacted with an interface character (hence, studying the lower part of the I-PEFiC model which concerns the character's affordances, see Fig. 1). In one study we investigated the role of aesthetics and affordances in the engagement process. Therefore, we designed beautiful and ugly looking interface characters within the Sims2 game environment with which users could actively interact (Van Vugt et al., 2006a). In addition, we designed a web-based application in which users interacted with an interface character in the role of a health advisor (Van Vugt et al., 2006b). With our studies, we aim to provide a more comprehensive understanding of human–character interactions.

In sum, understanding what determines user engagement is important because engagement highly predicts user satisfaction in human–character interaction. The present study showed that realism is not all that matters for user engagement with interface characters. User responses are also affected by the character's aesthetic appearance and its relevance to the task. The debate on the existence or non-existence of a persona effect may still be undetermined, but we now may suspect a more complicated picture in which various factors work together in predicting engagement, task performance and user satisfaction. An interface character is more likely to result in persona effects when it is engaging.

Acknowledgements

We are thankful to the Faculty of Social Sciences and the Faculty of Sciences, Vrije Universiteit in Amsterdam, who granted a 4-year research track (VUBIS) to the first author. We are grateful to Gerrit van der Veer and Jan Kleinnijenhuis for their generous support and fruitful discussions in the early stages of this project. Further, we would like to thank the students of the Multimedia Casus 2004 (spring) project and especially Ricardo Tangali for constructing the stimulus materials, and students of the Blaucapel College in Utrecht (The Netherlands) for their participation. Finally, we thank Herre van Oosten-dorp for his insightful comments on an earlier draft of this paper.

Appendix A. Recall test

The recall test is translated from Dutch.

Question 1. The feather is from a

- ostrich
- peacock
- pheasant
- heron

Question 2. Which color does the feather have?

- white
- gray
- light yellow
- light brown

Question 3. The length of the feather is

- seventy centimeters
- fifty centimeters
- forty centimeters
- twenty centimeters

Question 4. The stone is a

- marble stone
- glacier stone
- erratic boulder
- basalt stone

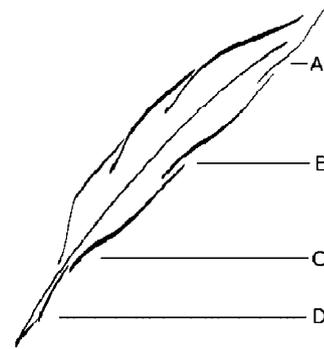
Question 5. The height of the stone is

- 13.5 cm
- 16.5 cm
- 9.5 cm
- 5.5 cm

Question 6. The artwork has to meet the following requirement:

- the feather is being kept in place by the weight of the stone
- the end of the feather is bent around the stone
- the stone is placed on the feather with its flat side
- the feather is bent after the placement of the stone

Question 7. Where is the stone placed on the feather?



Question 8. In the formation of the artwork the following requirement has to be met

- the feather should not reach beyond the column
- the feather has to reach beyond the column
- the feather has to be positioned as straight as possible
- the feather has to lie as flat as possible

Question 9. In the formation of the artwork, the following needs should be paid attention to:

- the amount of light in the space around the artwork
- the color of the light in the space around the artwork
- the angle under which the light is falling at the artwork
- the concentration of the light on the artwork

Question 10.

The column is

- rectangular
- hexagonal
- triangular
- cube-shaped

Appendix B. User perception questionnaire

Table 3 shows the reliability of each scale and the items of the user perception questionnaire.

Table 3
The reliability of each scale and the items of the user perception questionnaire

| Scale | Reliability (Cronbach's α) | Measures for no-character condition | Measures for interface character conditions |
|--------------|------------------------------------|---|---|
| Realism | .55 | n.a. | The figure looks natural The figure looks real The figure is a fantasy creature The figure looks fake |
| Aesthetics | .77 | The demo is beautiful The demo has a professional look The demo is ugly <i>The demo has an amateur look</i> | The figure is beautiful The figure has a professional look The figure is ugly <i>The figure has an amateur look</i> |
| Similarity | .82 | n.a. | I resemble the figure The figure and I are different |
| Relevance | .79 | What is said in the demo, is important I have learned something from the demo I do not care what is said in the demo If something is said in the demo, I think: never mind! | What the figure says, is important I have learned something from the figure I do not care what the figure says If the figure speaks, I think: never mind! |
| Valence | .90 | I can reproduce the artwork I can reproduce the artwork quickly I can reproduce the artwork without errors It is easy to reproduce the artwork The reproduction of the artwork will fail The reproduction of the artwork will take a long time I think I will reproduce the artwork wrongly It is difficult to reproduce the artwork | I can reproduce the artwork I can reproduce the artwork quickly I can reproduce the artwork without errors It is easy to reproduce the artwork The reproduction of the artwork will fail The reproduction of the artwork will take a long time I think I will reproduce the artwork wrongly It is difficult to reproduce the artwork |
| Involvement | .64 | n.a. | The figure is appealing The figure makes me happy |
| Distance | .78 | n.a. | The figure is annoying The figure irritates me |
| Satisfaction | .80 | The demo helped me to understand the information The demo was clear The demo gave exactly enough information The demo was confusing | The figure helped me to understand the information The figure was clear The figure gave exactly enough information The figure was confusing |

The item printed in *italics* was removed from the scale due to a bad fit. n.a., not applicable.

References

- Bailenson, J., Beall, A., Blascovich, J., Raimundo, M., Weisbuch, M., 2001. Intelligent agents who wear your face: User's reactions to the virtual self. In: *Intelligent Virtual Agents 2001*, pp. 86–99.
- Bailenson, J.N., Yee, N., Merget, D., Schroeder, R., in press. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *PRESENCE: Teleoperators and Virtual Environments*.
- Beun, R.-J., de Vos, E., Witteman, C., 2003. Embodied conversational agents: effects on memory performance and anthropomorphisation. In: Rist, T. et al. (Eds.), *IVA 2003, Lecture Notes in Computer Science 2792*, pp. 315–319.
- Brave, S., Nass, C., Hutchinson, K., 2005. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies* 62 (2), 161–178.
- Cassell, J., Thórisson, K.R., 1999. The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence* 13 (4–5), 519–538.
- Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., Stone, M., 1994. Animated conversation: rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In: *Proceedings of SIG-GRAPH 94*. ACM Press, pp. 413–420.
- Catrambone, R., Stasko, J., Xiao, J., 2002. Anthropomorphic agents as a user interface paradigm: experimental findings and a framework for research. In: *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Fairfax, VA, pp. 166–171.
- Catrambone, R., Stasko, J., Xiao, J., 2004. ECA as user interface paradigm. In: Ruttkey, Z., Pelachaud, C. (Eds.), *From Brows till Trust: Evaluating Embodied Conversational Agents*. Kluwer, pp. 239–270.
- Choi, Y.K., Miracle, G.E., Biocca, F., 2001. The effects of anthropomorphic agents on advertising effectiveness and the mediating role of presence. *Journal of Interactive Advertising* 2 (1).
- Davis, F.D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 13 (3), 319–339.
- De Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., De Carolis, B., 2003. From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies* 59 (1–2), 81–118.
- De Rosis, F., Cavalluzzi, A., Mazzotta, I., Novelli, N., 2005. Can embodied conversational agents induce empathy in users? In: *AISB'05 Virtual Social Characters Symposium*. Hatfield.
- Dehn, D.M., Van Mulken, S., 2000. The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies* 52 (1).
- Dillman, D.A., 2000. *Mail and Internet Surveys: The Tailored Design Method*, second ed. John Wiley and Sons, New York.
- Dryer, D.C., 1999. Getting personal with computers: how to design personalities for agents. *Applied Artificial Intelligence* 13 (3), 273–295.
- Eliens, A., Huang, Z., Visser, C., 2002. A platform for embodied conversational agents based on distributed logic programming. In: *AAMAS Workshop – Embodied Conversational Agents - Let's Specify and Evaluate Them*.

- Gelderman, M., 1998. The relation between user satisfaction, usage of information systems and performance. *Information And Management* 34 (1), 11–18.
- Gratch, J., Marsella, S., 2005. Some lessons for emotion psychology for the design of lifelike characters. *Journal of Applied Artificial Intelligence (special issue on Educational Agents – Beyond Virtual Tutors)* 19 (3–4), 215–233.
- Green, S., Salkind, N., 2005. Using SPSS for Windows and Macintosh. *Analyzing and Understanding Data*, fourth ed. Prentice-Hall, Upper Saddle River, NJ (ECAs).
- Guadagno, R.E., Blascovich, J., Bailenson, J.N., McCall, C., in press. Virtual humans and persuasion: the effects of agency and behavioral realism. *Media Psychology*.
- Hayes-Roth, B., 2003. What makes characters seem life-like? In: *Life-like Characters. Tools, Affective Functions and Applications*. Springer, Berlin, pp. 447–462.
- Hoorn, J.F., Konijn, E.A., 2003. Perceiving and experiencing fictional characters: an integrative account. *Japanese Psychological Research* 45 (4), 221–239.
- Hoorn, J.F., Eliëns, A., Huang, Z., Van Vugt, H.C., Konijn, E.A., Visser, C., 2004. Agents with character: evaluation of empathic agents in digital dossiers. In: *AAMAS '04 Workshop on Empathic Agents*.
- Hoorn, J.F., Konijn, E.A., Van Vliet, H.C., Van der Veer, G.C., in press. Requirements change: fears dictate the must haves; desires the won't haves. *Journal of Systems and Software*.
- Huang, Z., Eliëns, A., Visser, C., 2003. 'Is it within my reach?' – an agents perspective. In: Rist, T. et al. (Eds.), *IVA 2003. Lecture Notes in Computer Science* 2792, pp. 150–158.
- Johnston, V.S., Oliver-Rodriguez, J.C., 1997. Facial beauty and the late positive component of event-related potentials. *Journal of Sex Research* 34 (2), 188–198.
- Kanazawa, S., Kovarb, J.L., 2004. Why beautiful people are more intelligent. *Intelligence* 32 (3), 227–243.
- Koda, T., Maes, P., 1996. Agents with faces: the effect of personification of agents. In: *Proceedings of HCI'96*. London, UK, pp. 98–103.
- Konijn, E.A., Bushman, B.J., in press. World leaders as movie characters? Perceptions of J.W. Bush, T. Blair, O. Bin Laden, S. Hussein. *Media Psychology*.
- Konijn, E.A., Hoorn, J.F., 2005. Some like it bad. Testing a model for perceiving and experiencing fictional characters. *Media Psychology* 7 (2), 107–144.
- Lester, J., Converse, S., Kahler, S., Barlow, T., Stone, B., Bhogal, R., 1997. The persona effect: affective impact of animated pedagogical agents. In: *Proceedings of CHI '97*. Atlanta, pp. 359–366.
- Lindgaard, G., Dudek, C., 2003. What is this evasive beast we call user satisfaction? *Interacting with Computers* 15 (3), 429–452.
- Liu, Y., 2003. Engineering aesthetics and aesthetic ergonomics: theoretical foundations and a dual-process research methodology. *Ergonomics* 46 (13/14), 1273–1292.
- Moundridou, M., Virvou, M., 2002. Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning* 18 (3), 253–261.
- Nass, C., Steuer, J., Tauber, E.R., 1994. Computers are social actors. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Press, Boston, Massachusetts, United States, pp. 72–78.
- Nass, C., Moon, Y., Fogg, B.J., Reeves, B., Dryer, D.C., 1995. Can computer personalities be human personalities? *International Journal of Human-Computer Studies* 43 (2), 223–239.
- Norman, D.A., 2004. *Emotional Design*. Basic Books.
- Nowak, K., Biocca, F., 2003. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments* 12 (5), 481–494.
- Persson, P., 1999. *Agneta & Frida: a narrative experience of the web?* AAAI Fall Symposium on Narrative Intelligence.
- Picard, R.W., Klein, J., 2002. Computers that recognise and respond to user emotion: theoretical and practical implications. *Interacting with Computers* 14, 141–169.
- Reeves, B., Nass, C., 1996. *The Media Equation: How People Treat Computers, Televisions and New Media Like Real People and Places*. Cambridge University Press.
- Rickenberg, R., Reeves, B., 2000. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. *CHI Letters* 2 (1), 49–56.
- Ruttkey, Z., Dormann, C., Noot, H., 2004. ECAs on a common ground, a framework for design and evaluation. In: Ruttkey, Z., Pelachaud, C. (Eds.), *From Brows till Trust: Evaluating Embodied Conversational Agents*. Kluwer Academic Publishers, Dordrecht/Boston/London, pp. 27–66.
- Sproull, L., Subramani, R., Walker, J., Kiesler, S., Waters, K., 1996. When the interface is a face. *Human-Computer Interaction* 11 (2), 97–124.
- Stevens, J., 2002. *Applied Multivariate Statistics for the Social Sciences*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Takeuchi, A., Naito, T., 1995. Situated facial displays: towards social interaction. In: Katz, I., Mack, R., Marks, L., Rosson, M.B.J., Nielsen, E. (Eds.), *Human Factors in Computing Systems: CHI'95 Conference Proceedings*. ACM Press, New York, pp. 450–455.
- Tractinsky, N., 1997. Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Press, Atlanta, Georgia, United States, pp. 115–122.
- Tractinsky, N., Katz, A.S., Ikar, D., 2000. What is beautiful is usable. *Interacting with Computers* 13 (2), 127–145.
- Van Mulken, S., André, E., Müller, J., 1998. The persona effect: how substantial is it? In: *HCI '98: Proceedings of HCI on People and Computers XIII*. Springer-Verlag, London, UK, pp. 53–66.
- Van Vugt, H.C., Hoorn, J.F., Konijn, E.A., 2004. Adapting empathic agents to user experiences. In: *AAMAS '04 Workshop on Empathic Agents*. New York.
- Van Vugt, H.C., Hoorn, J.F., Konijn, E.A., de Bie Dimitriadou, A., 2006a. Affective affordances: improving interface character engagement through interaction. *International Journal of Human-Computer Studies* 64 (9), 874–888.
- Van Vugt, H.C., Veldhuis, J., Konijn, E.A., Hoorn, J.F., 2006b. Why fat interface characters are better e-health advisors. In: Gratch, J. et al. (Eds.), *IVA 2006. Lecture Notes in Computer Science* 4133, pp. 1–13.
- Venkatesh, V.G.M.M., Davis, G.B., Davis, F.D., 2003. User acceptance of information technology: toward a unified view. *MIS Quarterly* 27 (3), 425–478.
- Xiao, J., Catrambone, R., Stasko, J., 2003. Be quiet? evaluating proactive and reactive user interface assistants. In: *Proceedings of INTERACT '03*. Zurich, Switzerland, pp. 383–390.
- Xiao, J., Stasko, J., Catrambone, R., 2004. An empirical study of the effect of agent competence on user performance and perception. In: *Proceedings of AAMAS'04*. New York, pp. 178–185.