

Umeå University  
Cognitive Science Programme  
Master's thesis 20 p, spring 2003

# The Appearance of Character in Interaction with Screen-Based Computer Artifacts

Roger Andersson

Supervisors: Dr. Lars-Erik Janlert, Department of Computer Science  
Dr. Erik Stolterman, Department of Informatics



A general tendency in line with the technical development within modern information technology is that everything from personal devices to sophisticated information systems becomes increasingly more complicated to understand and to use. One approach to deal with this matter is put forward by Janlert and Stolterman (1997) who draws attention to the putative fact that people have the habit to think of artifacts as having character. According to Janlert and Stolterman, this works as a natural and cognitively efficient way to get a high-level understanding of an artifact, which they anticipate to be especially helpful and important in interaction with highly complicated artifacts. The observations and assumptions in the theoretical framework of Janlert and Stolterman have not been subject to systematic research or extensive exploration. This thesis takes the first steps to explore the notion of an ascription of characteristics within the area of screen-based computer artifacts through two experiments. These experiments are based on user evaluation of the characteristics ascribed to a set of stimuli computer programs through the judgment of semantic descriptors mainly originating in the domain of personality assessment. The results show that participants in general agree to use the descriptive concepts associated with the semantic descriptors to describe their impression of the stimuli programs. Analysis also shows significant differences between the impressions and ascriptions of characteristics with respect to different configurations of the stimuli programs that are used. The procedure is critically discussed and suggestions for further development and refinement of the procedure are pointed out. Also, future areas of investigation and further methodological approaches are presented as a means to achieve for a more comprehensive evaluation and understanding of the notion of the character of computer artifacts.



## *Introduction*

Progress in computer technology has led increasingly complex computer artifacts to be extensively used in today's society. In line with this development, computer artifacts are introduced in new contexts, new kinds of features are becoming available and new groups of people are using them in their daily lives. As a consequence, regular users of computer artifacts are in a higher extent facing situations that are complicated, demanding or go beyond what is familiar to them. Such downside of technological development is discussed by Norman (1998), who proposes that devices that simplify life at the same time make life harder by providing an increased complexity of use. Norman advice's that such troublesome situations of use can be supported or avoided through designing artifacts that provide information, or in his definition *mental aids*, that help the user to know what to do and how to behave in interaction. Laurel (1993), Suchman (1987) and Reeves and Nass (1996) among others have discussed and pointed out the resemblance between human-computer interaction and the social interaction of everyday life. In Suchman's perspective, people ascribe intentionality to and treat computer artifacts according to patterns of social interaction, due to a relative inability to get a look at the underlying processes of artifact behaviour. Attention to the abilities that are employed in everyday social situations may hence contribute to an understanding that is valuable to the field of human-computer interaction. Such attention then also offer a potential to more appropriately design mental aids in complicated artifacts, in Norman's sense. In social life, people make immediate conclusions and inferences on the basis of perceptual information and situational knowledge, despite no or little prior experience of the observed person (Jones, 1990; Fiske, 1991; Goffman, 1959). Such impression of a person is a conception of a whole - not strictly an account of perception of individual characteristics but rather an overall impression of the person's character (Asch, 1946; Asch & Zukier, 1984). An example of this attribution of character can be found within the theatre, where characters are ascribed to a role through distinct features of appearance and behaviour (Laurel, 1993; Janlert & Stolterman, 1997).

Janlert and Stolterman (1997) discuss the notion that people seem to relate and think of artifacts in their everyday lives similar to the objects of social life, thus making ascriptions of character to artifacts. According to Janlert and Stolterman, such an ascription works as a cognitively efficient way to get an understanding that guides interaction with an artifact. In their definition, the *character is an emergent unity of a number of characteristics displayed in an artifact*. They propose that characterizations may not be very precise and detailed, but rather, ascribing characteristics aims at making a higher-level, overall description that will cover several different aspects of an artifact. This kind of description will have a wide range of applicability that generates both general expectations and an understanding of artifacts' constraints. In Janlert and Stolterman's view, a description of characteristics may encompass *technical* and *functional* qualities, *aesthetical* qualities as well as *ethical* qualities (i.e. the artifact's social and moral relation towards its surrounding). More specifically, they suggest that to apply attribution of character to an artifact serves the purpose of: *generating expectations*, thus aiding in anticipating the behaviour; *generating explanations* of behaviour; as a *context for interpretation* of appearance and behaviour; and as a *schema for normal functioning*.

Janlert and Stolterman (1997) anticipate that relying on an overall feeling that guides interaction is of special importance with complex artifacts, whose range of behaviour, functionality and use is difficult to grasp. They discuss and identify two general implications for the design of computer artifacts, in that we should: (i) pay attention to the character and the completeness and coherence of its characteristics; (ii) develop a design of characteristic features to make the character of the artifact appear. To explain the importance of these implications, Janlert and Stolterman take cars as an example – it's essential that a car that goes fast and feels safe to drive, actually is safe and doesn't cause an excessively careless behaviour which risks somebody's life. On the other hand, it is also important that artifacts that are designed to be safe, and which technically satisfies this quality also appear safe and feel safe to use. Attention to and awareness of the ascribed characters of artifacts is thus of importance both for understanding the practice of interaction between people and artifacts but also of high relevance when complicated artifacts are to be designed to support easy and more effortless interaction. The approach proposed by Janlert and Stolterman (1997) differs from the both acclaimed and criticized notion of metaphors and analogies in a significant way. Both functional and non-functional metaphors and analogies relate different type of properties of one object to another object, as for example when drawing attention to the similarities between a word processor and a sports car, which is discussed by Janlert and Stolterman. But metaphors do not give the same sort of detailed information as an ascription of characteristics does, as they are more general and open to interpretation. When saying that the word processor resembles the sports car, this does not include information on the extent and boundaries of this resemblance.

Presently, systematic research with explicit use of the notion of character as defined by Janlert and Stolterman has not been conducted. However, studies such as Smets and Overbeeke (1995) verify that that it is possible to evaluate how a certain appearance better suits and explains a specific function by showing that people are able to discriminate between designs of packages created for different tastes of a dessert. The purpose of the present thesis is *to explore the assumption of an ascription of characteristics to computer artifacts*. The thesis aims *to establish a general procedure of empirical measurement and evaluation of characteristic attributes in computer artifacts that connects specific impressions of characteristics to specific design properties*. Such a procedure may be used as an essential element of a methodology for further investigation of the concept and for matters of practical design of appropriate characteristics in computer artifacts. This investigation starts with a review of methodological suggestions for the present study. The study itself is conducted through two experiments that investigate the patterns in the characteristics that are ascribed to a set of screen based computer programs created for the purpose of the study. In these programs, various aspects of its visual appearance, behavioural appearance and technical functionality are manipulated across the configurations that are used. The term *visual appearance* is henceforth used to replace the term *aesthetical quality* and is referring to visual screen elements such as shape and colour. The term defined as *ethical quality* by Janlert and Stolterman is henceforth replaced by the more straight forward term *behavioural appearance*. This behavioural appearance is visually presented on the screen but the term is used to refer to the emergent behaviour that originates in the course of interaction and

use. The programs are intentionally simple and of limited scope to gain high control of the properties of the programs. Using a set of simple programs also aims to investigate the assumption that ascription of character and characteristics is made easily even in settings where little information is to be received from the observed object, as been documented in research within the field of social interactions (Jones, 1990; Fiske, 1991; Asch, 1946; Asch & Zukier, 1984; Goffman, 1959).

#### *Review of implications for the practical procedure*

Janlert and Stolterman (1997) discuss industrial design as an area where symbolic qualities are explicitly used to express properties of objects, and where methods for designing such symbolic properties are established. The *product semantics* of industrial design investigates how physical forms and configurations of products work as *signs* of the product's properties that may transmit experiences and meanings to a beholder. According to Monö (1997) products can be considered as having technical, practical and semantic functions, the latter being: *to describe* the product's purpose, way of use and handling; *to express* values and qualities of the product; *to exhort* certain actions and reactions; *to identify* the product's origin and affiliation. The product semantic approach is applied through a single individual's subjective interpretation of the product's sign (Vihma, 1995). There is however little knowledge established concerning how various properties of screen-based artifacts influence people's impressions of them. In general, with the interpreter being a designer or design theorist this procedure runs the risk of a limited validity among general users of a product. The formalization of the semantic functions as discussed in Monö (1997) is however of interest for an extensive account of what impression a product gives. But, as the framework is basically limited to shapes, product semantics has little to say about behaviour and interaction.

The *semantic differential technique* (Osgood, Suci & Tannenbaum, 1957) was initially developed as an attempt to subject meaning to quantitative measurement. This technique makes it possible to investigate what meaning or impression people ascribe to a certain concept. The technique works through the rating of words affiliation with a concept (such as "Republican") on a contrasting bipolar scale. Each such scale can be represented as a linear function that passes through the origin of a semantic space (Osgood et al.). A sample of many semantic scales represents a multidimensional semantic space. As many of the scales are more or less overlapping in the semantic space they still may be efficiently and reliably defined when the scales are clustered to a lesser number of dimensions, which is conducted through a factorial analysis. The findings of Osgood et al. indicate that the semantic space of the meaning of concepts generally has numerous dimensions but that some of them repeatedly return. The recurring dimensions or "constituting factors" of correlated variables in their study were *evaluation*, *potency* and *activity*.

The decisive element of the semantic differential technique is the *factorial analysis* (FA) as introduced above. The FA is a statistical technique that finds linear combinations of variables that accounts for as much variation among the original variables as possible. When applied patterns of variations among observed variables are summarized so that a smaller number of factors can be used to replace the original variables (Tabachnik &

Fidell, 1996). Such factors are formed of coherent subsets with high internal correlation, which are relatively independent of other subsets. These factors are thus considered as reflecting underlying relations between the variables in the subset (Tabachnik & Fidell). The procedure of FA may briefly be described as an initial extraction of factors out of the observed variables, followed by a subsequent rotation that improves the interpretability and utility of the factors in the solution. There are several similar methods available for extraction and rotation of factors, among the most used is the Principal Components Analysis (PCA) for extraction and Varimax for rotation (Tabachnik & Fidell).

*Kansei Engineering* (KE) is a Japanese product development methodology, which uses the semantic differential technique to account for the user experiences and impressions of a product. The methodology exists in a variety of different types that can be used both to evaluate prototypes and finished designs, as well as to interpret a context of use in order to gain an insight in essential qualities needed in that context (Schütte, 2002). A KE-approach concerned with the evaluation of a design for a defined product domain is roughly implemented by the following procedure: At first a large pool of words, predominantly adjectives, which describe the product is collected. The following step is to have subjects evaluate a set of one or several designs on a semantic scale based on these words. Then the words are clustered into related components through FA. After clustering, suitable and representative words from each group are chosen as the “Kansei words” that are to be used in the following evaluation. Hence the number of words used in this second evaluation is reduced to a much lesser quantity in comparison to the initial pool of words. Although this results in less data material for analysis, it provides a more compact evaluation, which eases the workload and reduces the fatigue of the participants, according to Schütte. The Kansei words used in the second evaluation are connected to explicit product properties and design elements by analysis through linear regression. A connection how diverse properties influence an impression may thus be established (Schütte). The focus of KE is however to develop products which give impressions consistent with a brand, not on making products comprehensible and tractable. There is no concern for how the impression of a product may guide interaction, which is clearly emphasized within the character of things approach of Janlert and Stolterman (1997).

Research in social cognition considers the importance of how attributes ascribed to an observed person varies with regard to the information present in the social setting (Jones, 1990; Fiske, 1991). Methods of measurement of individual traits have been established through the research undertaken in the field of personality assessment. At present time, the Big-Five factor model of personality (Goldberg, 1993) appears as the most successful in outlining the basic components of personality. According to this model the space of personality attributes is spanned by five components, conventionally labelled *EXTRAVERSION*, *AGREEABLENESS*, *CONSCIENTIOUSNESS*, *EMOTIONAL STABILITY* and *INTELLIGENCE* (also named *OPENNESS TO EXPERIENCE*) (Goldberg, 1990). However, although the big-five factors produce the most stable patterns the specific number of factors is still not completely agreed on and seems to depend on the selection of the variables used for assessment (Saucier, 1997).

The rationale of the Big-Five model is the lexical hypothesis introduced by Francis Galton (Goldberg, 1993). This hypothesis states that the most important individual differences in human appearance and being may be coded in language as single terms. Years of research on this personality taxonomy have established the large inventory of words that are used as lexical *markers* of personality. These kinds of assessments are not conducted as an evaluation of specific individuals but to search for structure in the essential attributes, which account for the foundations of personalities. The procedure is, much simplified, to let a group of subjects evaluate a given person (as themselves or acquaintances) on a large number of markers. FA is used to investigate the structure in how these variables coincide with each other (Goldberg, 1990; 1992).

Sivik and Taft (1992) also apply a similar method of FA to explore people's associative impressions of the meanings of colour combinations. They find that every one of the five dimensions derived in their study (and different from the Big-Five factors) is associated with distinct colour combination attributes. They imply that concordance in judgment between people is high in this particular domain, although there are differences between groups of subjects.

### *Experiment 1*

The approaches reviewed above do not have the same focus and purpose as the character of things framework. As the semantic differential technique has been successfully combined with FA in various approaches to explore the subjective impressions within different contexts it however makes an input to the practical methodology. Crucial to the use of the semantic differential scale is the descriptive words selected for the semantic differentials. On account of the exploratory nature of the present study it is uncertain which set of words that best suit the present situation as descriptors of characteristics. For practical reasons, an investigation of a restricted range of descriptive words that does not put too extensive workload on the participants was considered as appropriate. This conveys a choice between different directions of the questionnaire. The descriptive words used in KE and on studying colour combinations in high extent refer to properties of visual and primarily physical appearance. It was considered that the descriptive words most adequate to represent the ascription of characteristics during *interaction* with an artifact are oriented more toward the domain of the social world and of describing individual's personality. An initial pool of semantic differentials was hence created out of the words used for investigating the Big-Five factor structure found in the inventory of Goldberg (1992). These words, termed as Big-Five *markers* present a concise assembly of words that still provide a robust reproduction of the Big-Five-factor structure. This choice of direction is thus not to be interpreted as an assumption that an artifact's character spans the same semantic space as personality characters but rather as a practical decision for the cause of methodological investigation. The impressions of an artifact's agency are also of interest to investigate for several reasons. As discussed above, this was shown to be an essential element in the studies of Osgood et al. (1957). Also, according to previous experience, the activity is a particular instance in people's attribution of qualities to artifacts. However, the Big-Five model notably ignores activity-related characteristics. Hence semantic differentials describing levels of activity were also included in the questionnaire, which construction is described more extensively below.

A classic study of Heider and Simmel (1944) shows the ease by which people ascribe meaning to simple geometrical figures appearing in an animated film. The geometrical figures were generally seen as animated beings and their functional behaviours were attributed as depending on motives and intentions. This use of simple forms with a spatial and temporal behaviour works as inspiration for the computer program that was constructed for this initial experiment. The purpose of this experiment is to *investigate and validate the ascription to and impression of characteristics in this simple program*. A procedure for assessment of the impressions and ascriptions of characteristics is applied *through an evaluation of contrasting semantic words in a questionnaire*, constructed for this purpose. The underlying structural relationships between these words are explored to investigate the regularity and concordance in the impressions among subjects. Specifically, this exploration may provide input to the construction of a comprehensive and compact tool for evaluation of characteristic properties, which still is powerful and adequate by including the most essential dimensions of the characters appearance. Apart from an evaluation through the questionnaire, the experiment also included a brief interview where subjects were asked about their impressions of the stimuli program and experiences of computer programs in general. This material is not included in the presentation of results but provided valuable input for further exploration.

### *Method*

The study was conducted using a questionnaire consisting of 31 semantic differential pairs for measurement of the subjects' ascriptions and impressions of characteristics in the stimuli program. The semantic pairs chosen for the questionnaire are intended to go beyond just describing properties as performance and effectiveness, which in general are in focus when evaluating computer artifacts. The construction of the questionnaire is described in detail below. Subjects' ratings of the 31 pairs on this 7-grade scale are the dependent variables of the experiment. Independent variables of the experiment are the configurational properties of the stimuli program as described below.

### *Participants*

12 subjects, 8 males and 4 females, aged between 20 and 28 years where recruited for this study. Refreshments were offered as compensation for participation. All subjects were undergraduate students in humanities or social sciences with Swedish as their native language. Subjects participated voluntarily and were assigned three out of eight configurations, described below. Each subject thus was part of three trials, where the configurations where ordered randomly between subjects. The total time to perform the experiment varied between different subjects, ranging from 30 to 60 minutes (due to the interview).

### *Material*

A simple computer program was created in Macromedia Flash MX. The program, henceforth called the *stimuli program* produces repeatedly animated events with objects that depend on user actions. Three properties of the program systematically vary in two ways. This results in a 2x2x2 factorial between-groups design, with a total of eight different configurations of the same program. The properties of variation in the program

are *colour*, *object form* and *modes of interaction*. The intent is not primarily to explore what specific properties elicits a characteristic to be ascribed to a program, but rather to establish a method able to distinguish a relationship between properties and characteristics. These configurations may to some extent be interpreted as including a variation of more than one property within a configuration. This is deliberate and intends to make the configurations as different as possible with respect to the given program to elicit well separated impressions of the configurations. The set of *colours* are: a warm **red** (RGB: #FF3333); a cold **blue** (RGB: #00CCFF). The set of *object forms* are: a transparent, twinkling **triangle** with a soft movement; a larger, solid **square** with a jagged movement. The set of *modes of interaction* are: an **unambiguous mode** with two square forms functioning as buttons, situated on separate sides of the lower part of the screen that is controlled through clicking - one for adding animated objects in a neat arrangement, one for deleting animated objects; an **ambiguous mode** with four square forms functioning as buttons, tightly grouped on the lower right part of the screen, that respond both to clicking and dragging the mouse over them - three different buttons add animated objects with random rotation and position, and one button delete these animated objects. The configurational properties were labelled according to their initial letter, consequently the label is R for **red**, B for **blue**, T for **triangle**, S for **square**, U for **unambiguous mode** and A for **ambiguous mode**. The configurations are labelled through the combination of what configurational properties they possess. Table 1 shows these labels and the number of subjects assigned to each configuration. These labels, descriptions and characterisations presented within the text are not shared with the subjects, but are used as a point of reference for own usage and for the presentation. It is possible that participants would not agree on some of the descriptions and characterisations of the different configurations.

Table 1: Labels for configurations and configurational properties. The amount of subjects within each configuration is shown within parenthesis. Abbreviations: R for **red** and B for **blue**; T for **triangle** and S for **square**; U for **unambiguous mode** and A for **ambiguous mode**.

	Unambiguous mode		Ambiguous mode	
	Square	Triangle	Square	Triangle
<b>Red</b>	<i>RSU</i> (5)	<i>RTU</i> (4)	<i>RSA</i> (4)	<i>RTA</i> (5)
<b>Blue</b>	<i>BSU</i> (4)	<i>BTU</i> (4)	<i>BSA</i> (4)	<i>BTA</i> (6)

The stimuli program was dimensioned 800x600 pixels and was displayed as a Flash movie in Microsoft Internet Explorer on a standard PC with a 17" screen with a resolution of 1280x1240 pixels. The Internet Explorer had been configured to hide all superfluous toolbars, hence only the "File bar"-menu was visible. Figure 1 and 2 below shows how two of the configurations of the stimuli program may look at use.

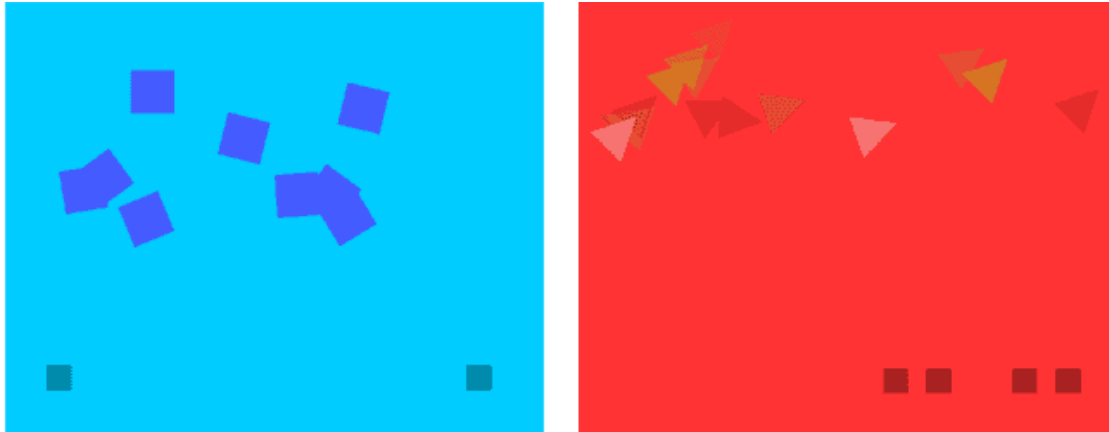


Figure 1 (left) and 2 (right): Example of the look of the configurations during trials. To the left is configuration *BSU*, to the right is configuration *RTA*.

*Construction of the questionnaire*

Among the 100 markers of traits provided by Goldberg (1992), there are twenty markers loading specifically on each of the Big-Five factors (*EXTRAVERSION*, *AGREEABLENESS*, *CONSCIENTIOUSNESS*, *EMOTIONAL STABILITY* and *INTELLIGENCE*). In Goldberg's inventory, the words functioning as markers are classified as being of positive or negative valence. The inventory is composed of ten markers each of positively and negatively valence, with the exception of the words belonging to the factor *EMOTIONAL STABILITY*. It is hard to find markers of positive valence for this factor why the negatively valued words are over-represented in the inventory (Goldberg). Markers considered as fitting for the present experiment were translated into Swedish using a standard dictionary (i.e. Petti, 2000). In the translation there are some cases several synonymous Swedish words that correspond to a single English word. In choosing between the synonyms, the primary criteria was the ordering of the words in the dictionary, but the meaning and clarity when using a word within the present domain was also considered. All translated Swedish words were collected in a list and grouped in a category corresponding with the original Big-Five factors among the English markers. Appropriate words were also added from the inventory of Saucier (1997) that provided a larger selection when the words in Goldberg (1992) were considered as insufficient.

Five positive and five negative words from each category were selected out of the collected list of Swedish words for use in the questionnaire. The collected Swedish words were selected by considering their meaning and apparent appropriateness for describing the characteristics of the present program domain. For example, it was desirable that the markers were not considered too specifically human in meaning (as for example *verbal*), that they not too clearly signified performance (as given by the discussion above). These selected words were paired two and two as a positive-negative contrasting pair on a bipolar scale. The use of contrasting pairs was intended to more clearly specify the meaning of each word on the scale, to avoid different interpretation by different subjects. Due to the interest in the impression of activity as discussed above, another twelve words (six of positive value and six of negative value) that were considered as good descriptors of activity were also paired as contrasts. As all words selected from the Big-Five

inventory were adjectives, these words were also selected among adjectives. This procedure resulted in a total selection of 31 bipolar pairs for the questionnaire, distributed over six different categories of origin. Five pairs originate in each one of the Big-Five factors respectively and the sixth pair originates in the activity category. This is illustrated by Table 2.

Table 2: Origin for categories with examples of words in each category.

Category	<i>EXTRA- VERSION</i>	<i>AGREE- ABLENESS</i>	<i>CONSCIENT- IOUSNESS</i>	<i>EMOTIONAL STABILITY</i>	<i>INTELLIGENCE</i>	<i>ACTIVITY</i>
Example of words	Assertive, Introverted	Pleasant, Cold	Stable, Careless	Calm, Tense	Intelligent, Naive	Active, Passive

The 31 pairs of words were mixed as the questionnaire was compiled, so that pairs representing the six different categories were scattered and not adjacent. The order of the positively and negatively valued word in each pair was also varied so that words of positive and negative value were equally often in the left and right position on the bipolar rating scale. The pairs were assigned an index between 1 and 31 in accordance to their position in the questionnaire. The same order was used for all completions of the questionnaire. The rating scale was graded in seven discrete points. Each grade was associated with one of the descriptive concepts (translated from Swedish): *Extremely*, *Greatly*, *Slightly* and *Neither*, as shown by figure 3. The questionnaire together with instructions appears in Appendix A (in Swedish).

Extremely Greatly Slightly Neither Slightly Greatly Extremely

0. Powerful: \_\_\_\_\_: \_\_\_\_\_: \_\_\_\_\_: \_\_\_\_\_: \_\_\_\_\_: \_\_\_\_\_: \_\_\_\_\_: Powerless

Figure 3: The 7-grade rating scale of the questionnaire, with an example of a contrasting pair (this actual pair was used as an instructive example in the questionnaire). The "0" to the left shows the index number of the pair. The descriptive concepts are shown at the top.

Below, when pairs or descriptors from the questionnaire in study are presented in the text an English translation of the Swedish word is used. For convenience, positively valued words are presented first even if this is not the case in the questionnaire.

#### *Procedure*

The experiments were conducted in a secluded room at the Department of Computing Science at the University of Umeå. As introduction, subjects were reviewed about the procedure of the experiment, telling them they would be acquainted with three computer programs, which they were expected to explore by themselves. No descriptions were given about the functionality of the program or the program itself. They were also informed they were expected to later answer questions about the programs. The subjects were instructed to themselves decide when they were done with each program. The experiment continued with the presentation and exploration of the first configuration that

had been assigned to the subject. After the first trial, subjects were given the questionnaire together with its written instructions. These instructions describe the study, and informs on how to answer the questionnaires by selecting the descriptive concept that best correspond to their impression of the program on each pair. After this, subjects completed the questionnaire, which took approximately 5 minutes. This process was repeated for the second and third trials: an exploration of an assigned configuration and then the completion of the questionnaire belonging to each configuration respectively. After the completion of the third questionnaire, the subjects were instructed to indicate what pairs of contrasting words they considered as inappropriate for describing their impression of the last configuration of the stimuli program. The experiment finished with a second interview, discussing what type of ascriptions the subject experienced he or she had made to the stimuli program, and what ascriptions he or she make to other computer programs in daily usage.

### Results

The number of pairs considered as inappropriate range between 0 and 19 pairs (mean: 6.9) for different subjects. The number of subjects considering individual pairs as inappropriate range between 0 and 7 (mean: 2.7) for different pairs of words respectively. These results are depicted in Table 3.

Table 3: Column I show subjects' individual patterns of selecting pairs as inappropriate; Column II shows the overall pattern of selecting individual pairs as inappropriate.

	Column I. Subject's individual pattern of selecting pairs (Total pairs = 31)	Column II. Pattern of selecting individual pairs (Total subjects = 12)
<i>Mean</i>	6.917	2.678
<i>Median</i>	7	2
<i>Std. dev</i>	5.107	2.040
<i>Min selected</i>	0	0
<i>Max selected</i>	19	7

The pair VITAL-RESIGNED was considered as inappropriate as describing the stimuli program by more than 50 % of the subjects (7 out of 12). The pair's BOLD-UNOBTRUSIVE, MERRY-INTROVERTED and WISE-NAIVE considered as inappropriate by exactly 50 % (6 out of 12) of the subjects, whereas the rest of the pairs were below this 50 % criteria. Four of the pairs, WARM-COLD, RELIABLE-UNRELIABLE, ACTIVE-PASSIVE and EFFECTIVE-IDLE were not selected by any subject as inappropriate.

All ratings of the pairs on the 7-graded scale were completed by the subjects. This resulted in no missing data points among the dependent variables and a total of 36 complete questionnaires. Subjects' rating on the dependent variables was ranked between -3 and +3 in seven discrete steps. Consequently, judging a pair through selecting the descriptive concept *Extremely* for the negatively valued word the corresponding variable

was ranked as -3. A judgment of *Extremely* for the positively valued word the corresponding variable was ranked as +3. A judgment of *Greatly* was ranked as -2 or +2, *Slightly* as -1 or +1 and *Neither* as 0.

The sample's normality distribution was derived from the ratings on the dependent variables with respect to its skewness and kurtosis. 4 out of 31 variables passed the Kolmogorov-Smirnov test for normal distribution, with the conventional significance level of .01. These variables were STABLE-SLOPPY, PERFORMING-LACKING INITIATIVE, MERRY-INTROVERTED and ENTERPRISING-UNENTERPRISING. A FA with PCA extraction and Varimax rotation was hereafter conducted including all variables for all configurations. The extraction with Varimax rotation resulted in seven components with eigenvalues higher than 1.0 that account for over 80 % of the total variance in the solution. Table 4 presents the seven components converged in 9 iterations with a Varimax rotation. Variable loadings greater than 0.71, with 50 % overlapping variance are considered as *excellent*. Variable loadings greater than .63 (with 40 % overlapping variance) are considered as *very good*, greater than .55 (30 %) as *good*, greater than .45 (20 %) are considered as *fair* and .32 (10 %) are considered as *poor* (Tabachnik & Fidell, 1996).

The component structure derived from this extraction shows that five of the variables have their primary loadings on the first component. Loadings on this component range between .94 and .74. Examples of pairs coinciding with this first component are RELIABLE-UNRELIABLE and STEADY-SLOPPY. The second extracted component includes seven pairs of primary loadings between .87 and .50. For example, pairs loading on the second component include ACTIVE-PASSIVE and WARM-COLD. The extracted third component includes four pairs, among these are MERRY-INTROVERTED. Loadings for primary variables on the third component are ranging between .82 and .69. Primary loadings on the fourth component range between .86 and .50, and include five pairs; among these is RELAXED-NERVOUS. Primary loadings on the fifth component range between .84 and .48, and for example include a pair as DEEP-SUPERFICIAL. Primary loadings on the sixth component come from two variables, ELATED-RESTRAINED and VITAL-RESIGNED with component loadings .85 and .62 respectively. The seventh component had primary loadings of .53 and .52 from the pairs BOLD-UNOBTRUSIVE and RELIEVED-TROUBLESOME.

The seventh component was excluded from further use as it had much lower primary component loadings from its constituting variables, which also had high secondary loadings (between .49 and .34) on four other components. Loadings further than the primary loadings are in general much lower than the primary loadings on all components, with the exception of the sixth and seventh components. The difference between the loadings on the primary component and the other components are much less legible for the last two components with the exception of the pair ELATED-RESTRAINED.

Table 4: Varimax-rotated component loadings derived in 9 iterations through PCA. Values equal to or larger than  $|\cdot| \geq .32$  are listed in boldface type. Category of origin and index in the questionnaire are listed together with the name of the pair. Original category name abbreviations: *AGR* for *AGREEABLENESS*; *CON* for *CONSCIENTIOUSNESS*; *ACT* for *ACTIVITY*; *EXT* for *EXTRAVERSION*; *INT* for *INTELLIGENCE*; *EMO* for *EMOTIONAL STABILITY*.

Pair	Category	Index	Component						
			1	2	3	4	5	6	7
RELIABLE–UNRELIABLE	<i>AGR</i>	20	<b>.94</b>	-.01	.16	.09	.06	.08	-.10
HONEST–ERRATIC	<i>CON</i>	21	<b>.92</b>	-.05	.20	.17	.08	.01	.00
STEADY–SLOPPY	<i>CON</i>	3	<b>.82</b>	-.01	.05	.34	.00	-.05	.07
CAREFUL–DISORDERED	<i>CON</i>	14	<b>.77</b>	-.09	<b>.49</b>	.04	-.07	-.21	-.02
DETERMINED–UNCERTAIN	<i>EXT</i>	12	<b>.74</b>	-.13	<b>.44</b>	.01	.14	-.05	.19
PERFORMING–LACKING INITIATIVE	<i>ACT</i>	31	-.17	<b>.87</b>	.03	.04	.28	.18	-.02
ACTIVE–PASSIVE	<i>ACT</i>	6	.23	<b>.86</b>	.16	-.21	.03	-.02	.08
EFFECTIVE–IDLE	<i>ACT</i>	11	.07	<b>.82</b>	.13	-.15	.01	-.10	.03
ENERGETIC–APATHETIC	<i>ACT</i>	30	<b>-.33</b>	<b>.81</b>	.21	.13	.24	.14	.00
WARM–COLD	<i>AGR</i>	2	.11	<b>.75</b>	.04	.24	.06	.42	.06
ENTERPRISING–UNENTERPRISING	<i>ACT</i>	18	<b>-.42</b>	<b>.66</b>	.23	.06	.21	.22	.09
PLEASANT–UNPLEASANT	<i>AGR</i>	26	.21	<b>.50</b>	.25	<b>.38</b>	.24	<b>.48</b>	.29
MERRY–INTROVERTED	<i>EXT</i>	19	.10	.18	<b>.82</b>	.08	-.01	.17	-.06
EXTROVERTED–INHIBITED	<i>EXT</i>	1	.24	.34	<b>.80</b>	.10	.04	.11	-.18
ACCOMMODATING–DEMANDING	<i>AGR</i>	8	<b>.41</b>	.16	<b>.73</b>	.19	-.09	.07	.25
CONSIDERATE–NONCHALANT	<i>CON</i>	9	<b>.36</b>	.02	<b>.69</b>	.27	-.05	.07	.25
RELAXED–NERVOUS	<i>EMO</i>	4	.16	-.05	.18	<b>.86</b>	.00	.04	.08
EASED–TENSE	<i>EMO</i>	28	-.05	.05	.10	<b>.80</b>	.16	-.06	-.16
SERENE–IMPATIENT	<i>EMO</i>	16	<b>.34</b>	-.13	.02	<b>.75</b>	.24	-.17	.00
OBLIGING–DISOBLIGING	<i>AGR</i>	14	<b>.34</b>	.29	<b>.47</b>	<b>.55</b>	.05	.21	<b>.32</b>
CALM–IRRITABLE	<i>EMO</i>	10	<b>.47</b>	.00	.28	<b>.50</b>	.22	-.04	.28
DEEP–SUPERFICIAL	<i>INT</i>	23	-.19	.00	-.16	.04	<b>.84</b>	.14	.13
COMPETENT–IGNORANT	<i>INT</i>	29	.10	.20	.01	.18	<b>.83</b>	-.03	-.22
INTELLIGENT–UNINTELLIGENT	<i>INT</i>	17	<b>.39</b>	.26	.11	.19	<b>.69</b>	.21	.06
REFINED–UNSOPHISTICATED	<i>INT</i>	11	-.07	<b>.40</b>	-.04	.12	<b>.68</b>	-.26	.30
WISE–NAIVE	<i>INT</i>	5	<b>.41</b>	.30	.17	-.03	<b>.57</b>	.20	-.13
CAUTIOUS–NEGLIGENT	<i>CON</i>	27	<b>.32</b>	-.13	.01	<b>.44</b>	<b>.48</b>	.18	<b>.33</b>
ELATED–RESTRAINED	<i>EXT</i>	25	-.13	.13	.12	-.15	.07	<b>.85</b>	-.12
VITAL–RESIGNED	<i>ACT</i>	24	.00	<b>.33</b>	<b>.40</b>	.04	.16	<b>.62</b>	<b>.39</b>
BOLD–UNOBTRUSIVE	<i>EXT</i>	7	-.21	<b>.49</b>	-.05	<b>-.35</b>	.14	-.02	<b>.53</b>
RELIEVED–TROUBLESOME	<i>EMO</i>	22	<b>.46</b>	.15	<b>.34</b>	<b>.37</b>	-.06	-.03	<b>.52</b>

Table 4 also show that while some of the original categories of the questionnaire are preserved within this extracted component structure there are others that are not preserved. The pairs belonging to the categories *ACTIVITY*, *EMOTIONAL STABILITY* and *INTELLIGENCE* have their primary loadings on the second, fourth and fifth components with the exception of one pair for each category respectively. Three out of five pairs of the *CONSCIENTIOUSNESS* category have their primary loadings on the first component. The pairs of the other two original categories *AGREEABLENESS* and

*EXTRAVERSION* are more scattered over the different components in a less systematic fashion.

Representative words were then selected as descriptors of each extracted component, according to their loadings on the component and due to their appropriateness as descriptors. The first extracted component was hence classified as “Reliability”, the second component as “Activity”, the third component as “Merriness”, the fourth component as “Relaxation”, the fifth component as “Competence” and the sixth component as “Elatedness”. These component classifications are represented in Table 5.

Table 5: Component classifications for the first study with pairs represented within each component.

Component classification	Representing pairs
“Reliability”	RELIABLE–UNRELIABLE, HONEST–ERRATIC, STEADY–SLOPPY, CAREFUL–DISORDERED, DETERMINED–UNCERTAIN
“Activity”	PERFORMING–LACKING INITIATIVE, ACTIVE–PASSIVE, EFFECTIVE–IDLE, ENERGETIC–APATHETIC, WARM–COLD, ENTERPRISING–UNENTERPRISING, PLEASANT–UNPLEASANT
“Merriness”	MERRY–INTROVERTED, EXTROVERTED–INHIBITED, ACCOMMODATING–DEMANDING, CONSIDERATE–NONCHALANT
“Relaxation”	RELAXED–NERVOUS, EASED–TENSE, SERENE–IMPATIENT, OBLIGING–DISOBLIGING, CALM–IRRITABLE
“Competence”	DEEP–SUPERFICIAL, COMPETENT–IGNORANT, INTELLIGENT–UNINTELLIGENT, REFINED–UNSOPHISTICATED, WISE–NAIVE, CAUTIOUS–NEGLIGENT
“Elatedness”	ELATED–RESTRAINED, VITAL–RESIGNED

To estimate the rankings on the components between different configurations as if they had been evaluated directly as single items of evaluation, the mean scores for the extracted components were estimated through the regression approach (described in Tabachnik & Fidell, 1996). The distribution of each estimated component score has a mean of zero and a standard deviation of 1. When ratings on each component are estimated by this procedure, their format may be hard to interpret manually but they nevertheless provide excellent material for statistical evaluation that takes account of systematic differences in the structure of the estimated component scores.

These estimated component scores were analysed through a linear regression model to explore differences in subjects’ impressions of the program according to its configurations. An analysis of variance between groups (ANOVA) shows there is a significant difference ( $p=.004$ ) between the configurations on the impression of the component “Reliability”. No other significant differences are found between the eight configurations. These results are presented in Table 6.

Table 6: ANOVA showing differences in the ascription of characteristics to the program, according to the configurations (N=36). Total SS for each component respectively is 35.0. Significant differences ( $p < 0.05$ ) are listed in boldface type.

	“Reliability”	“Activity”	“Merriness”	“Relaxation”	“Competence”	“Elatedness”
F	3.93	1.54	1.61	.95	1.39	1.418
Sig.	<b>.004</b>	.194	.173	.486	.249	.237

The estimated mean scores for the component “Reliability” are presented in Figure 3, which shows how the mean scores for the component “Reliability” are lower for all configurations with ambiguous mode of interaction. The biggest difference in F scores is between the configurations *BTA* (Mean = -.896) and *RSU* (Mean = 1.054) with a total SS of 35.0 and a maximal mean value of +3 and a minimal mean value of -3. The significance in the difference between these two means is  $p=.063$ , thus statistically significant only at a very moderate p-level.

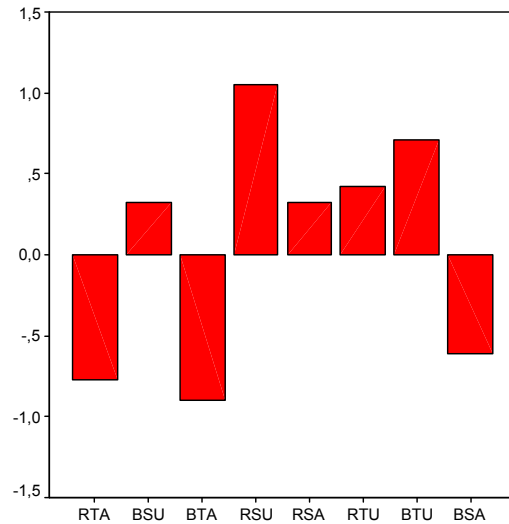


Figure 3: Estimated mean scores for the component “Reliability” due to configuration. Abbreviations: R for **red** and B for **blue**; T for **triangle** and S for **square**; U for **unambiguous mode** and A for **ambiguous mode**. Maximal score = 3, minimal score = -3.

A multivariate general linear model provides regression analysis and analysis of variance that show differences in the impressions according to the set of the three configurationally properties *colour*, *object form* and *mode of interaction* and also effects of interaction between these sets of properties. This renders the possibility to associate specific configurational properties to specific impressions. The results from this analysis indicate a significant difference in the impression of the stimuli program between the set of configurational properties of *colour* on the component “Activity” ( $p=.003$ ). The configurational property **red** is considered as more “active” in comparison with the property **blue**. Furthermore, the analysis indicates a significant difference in the impression of the stimuli program between the set of configurational properties of *object form* on the component “Relaxation” ( $p=.021$ ) and the *modes of interaction* on the component “Reliability” ( $p=.000$ ). The **triangle** is considered as more “relaxed” in comparison to the **square** whereas the **unambiguous mode** of interaction is considered as more “reliable” in comparison to the **ambiguous mode** of interaction. No other significant differences and no interaction effects are found between the set of configurational properties.

### *Discussion*

Results from the study indicate that subjects to a great extent are willing to ascribe properties that may be associated with the lexical descriptors used in the questionnaire. Among 31 pairs only 1 pair (or 3.2 %) was considered as inappropriate by more than 50 % of the subjects, which is a commonly adopted criterion. Furthermore, subjects' judgment on the inappropriateness of the pairs is not systematic, but scattered. Few words receive high or low ratings of inappropriateness and most pairs are selected as inappropriate by about 2 to 3 subjects (indicated by the mean and median values, 2.7 and 3 respectively). The pattern of describing the impression of the stimuli program varies between subjects; one subject considered no pairs as inappropriate, whereas another subject in the experiment considered 19 of the pairs as inappropriate. There is not one obvious explanation to these results; but the causes might for example be originating in previous experience and expertise, as discussed by Janlert and Stolterman (1997). The impression of the stimuli programs may also be influenced by a variety of other factors that are not accounted for within the experimental situation, such as: other previous references to the domain; differences in the interpretation of the semantic meanings of the pairs or single words; the time spent on the configuration; or strictly personal characteristics such as motivation and willingness to familiarize with the stimuli program. To sum up, subject's judgment of the pairs verifies that the characteristic traits addressed by the questionnaire may be ascribed to screen-based computer artifacts but that some of the words may be of limited utility for the actual domain of this experiment. There are moreover uncertainties regarding under what circumstances and conditions these ascriptions are most likely to be applied and how they influence the course of interaction.

The FA resulted in seven components, which accounts for most of the variance within the collected data. As the primary loadings on the seventh component were less distinct this component was excluded from further usage in the evaluation. These remaining six components did however not reproduce the original Big-Five categories. There are several possible explanations for this outcome. First of all, an analysis of 36 questionnaires is unfortunately not fully satisfying for a FA. Words in the big-five factor inventory also share a not insignificant correlation and meaning with other factors than its primary one (Goldberg, 1992). Consequently, it's not trivial to find words that exclusively load on only one factor and this may have impacted the results. Another possible explanation is that the dimension of the semantic space is not equal between different domains of use; hence the significance of words may change according to the context, as pointed out in the studies of Asch (1946). It is therefore conceivable that the content of a descriptor may change according to what entity it is used to evaluate and describe. In this case, the underlying structures of the semantic spaces of for example personality assessment and for describing screen based computer artifacts may be different, which was initially assumed. The extracted components as presented in Table 4 reveal some interesting correlations. An example is the correlation between pairs of words from the original category *ACTIVITY* and the pairs *WARM-COLD* and *PLEASANT-UNPLEASANT* that constitute the second component extracted in the FA.

Analysing the estimated component scores with the ANOVA show there is a significant difference in subjects' ascription of "Reliability" due to configuration of the stimuli

program. This difference may be derived from the *mode of interaction*, which apparently influence subjects' ascription of "Reliability" as shown by the general linear model analysis according to configurational properties. The results of the experiment show that the **unambiguous mode** appears as significantly more "reliable" and "stable" to the group of subjects in comparison to the **ambiguous mode**. Also indicated by the general linear model analysis is that *colour* influences subjects' impression of the "Activity" of the stimuli program. The **red** configuration appears as significantly more "active" than the **blue** configuration to the group of subjects. The *object form* influences the subjects' impression of the "Relaxation" of the stimuli program. The **triangle** appears as significantly "more relaxed" than the **square**. This difference according to *object form* may originate in the softer movement of the **triangle** in comparison with the jagged movement of the **square**.

Most variables in the sample were not normal distributed, and the FA was conducted without discarding these variables. The ANOVA statistic is robust to moderate departures from the normality assumption when sample sizes are equal or nearly equal (which is the case for a comparison between these configurations). The main concern for the statistical analysis is thus that the factor extraction may be lacking some statistical validity because of the variables that are not normal distributed. Moreover, the component extraction was conducted on the full sample of variables and not with respect to individual configurations due to the limited number of subjects in each single configuration that was not sufficient for a set of separate FA's.

An essential aspect of the stimuli program of this initial experiment is that the variations among configurations of the stimuli program mainly concern the visual appearance, as *colour* and *object form*. One property of the program, *mode of interaction*, was different between the configurations both according to the visual appearance, behavioural appearance and technical functionality. There are thus several important implications for further studies. With respect to the theoretical framework underlying this thesis (i.e. Janlert & Stolterman, 1997), recognizing behavioural appearances is of special interest as it is generally dismissed in investigations of screen-based computer artifacts as well as artifacts in general. Moreover, it is also of significance to look at the reproduction of the patterns of the extracted component loadings, as it within this experiment is indicated that the Big-Five category descriptors do not automatically reproduce the same components when they are used to evaluate objects different to the domain of personality. To explore these matters, a second experiment was conducted.

## *Experiment 2*

The initial experiment indicated the possibility to capture subjects' ascriptions to and impressions of characteristics in simple computer programs through the application of semantic differential scales. As the preceding experiment encompassed few participants the validity of the findings were hence limited. A second experiment with a larger number of subjects is conducted to attain a better validity and to further explore the implications discussed above. The stimuli program used in the second experiment intends to offer the participants possibilities to interact with the program to a greater extent. As discussed above, it is of special interest to recognize how characteristics may be ascribed

to a program with respect its behavioural appearance. The stimuli program for the second experiment hence primarily aims to shift between configurational properties of behavioural appearance and functionality that affects the course of interaction rather than the visual appearance of shapes and colours. By adopting practically the same semantic differential pairs for evaluation as in the initial experiment, it is possible to investigate the robustness and reproducibility in the extraction of components, which is of valuable input to an exploration of the methodology. The stimuli programs for assessment are however different in the two experiments. For this reason it must be pointed out that the main outcome is not a comparison of the characteristics ascribed to the particular programs, but rather to distinguish between the differences among the configurations within each experiment. This second experiment is intended both to *apply the methodology used in the initial study more extensively* but also to *examine the extraction of components within a slightly different type of domain to further investigate some of the implications of the initial experiment in a slightly different type of program*. For the purpose of the experiment a simple computer game is constructed, which comprise the task of collecting crosses with a vessel in a surrounding with obstacles. The rationale of playing the game is not intended to explicitly deal with maximal performance, but to encourage participants' curiosity and intrinsic motivation in the exploration and use of the program.

### *Method*

Starting with the questionnaire of the first study, another eight contrasting pairs were added to further elaborate the extraction of components. The rationale behind adding these words were that they provided an interesting addition to the words already present in the questionnaire by describing aspects of the current context that were not accounted for by the pairs of the initial experiment. The purpose is also to elaborate with how the construction of the questionnaire may affect the outcome of the FA. In the evaluation of appropriateness in the initial experiment, the pair VITAL-RESIGNED was considered as inappropriate by 7 out of 12 subjects. This pair was kept in the questionnaire of the second experiment, as this experiment concerned for a different domain of investigation. This resulted in a questionnaire of totally 39 pairs of words. Two of the words used in the first questionnaire were replaced with synonymous words considered as more suitable descriptors in the context of computer artifacts. The pair REFINED-UNSOPHISTICATED was replaced by REFINED-DULL and DEEP-SUPERFICIAL was replaced by SENSIBLE-SUPERFICIAL. Of the eight new pairs, four were selected from the inventories of Big-Five-factor markers in Goldberg (1990; 1992) and Saucier (1997). Out of these four pairs, three originated in the Big-Five category *EXTRAVERSION* (PLAYFUL-PRUDENT, VIGOROUS-COMPLIANT and ENTHUSIASTIC-BASHFUL) whereas the last pair originated in the category *CONSCIENTIOUSNESS* (INDULGENT-DOMINANT). The last four pairs, SAFE-UNSAFE, SECURE-INSECURE, HARMONIOUS-CHAOTIC and LIVELY-RESTRAINED were selected and paired out of Sivik et al's (1992) inventory and are categorized as *ADDITIONAL*. The new pairs were mixed up with the old pairs similar to the procedure of the construction of the first questionnaire, thus the pairs of the different origins were scattered and not adjacent. The positively and negatively valued words were also mixed within each pair, so that words of positive and negative value were equally ordered to the left and right of each other on the bipolar rating scale. The pairs were assigned an index between 1 and 39 in accordance to their position in the questionnaire. The same order

was used for all completions of the questionnaire. Subjects' ratings of the 39 pairs on this 7-grade scale are the dependent variables of the experiment. Independent variables are the configurational properties of the stimuli program as described below. The study was split up in two parts, a pilot study and a conclusive study. No data is presented from the pilot study, as it was carried out to refine the experimental procedure and to try out and calibrate the settings of the stimuli program. This questionnaire and its instructions appear as Appendix B (in Swedish).

#### *Participants*

75 participants between 20 and 33 years old were recruited for this study. Participation was voluntary and refreshments were offered as compensation. In the pilot study 11 subjects took part. The conclusive study consisted of 64 subjects, of which 32 were female and 32 were male. The subjects of both studies were either graduated or undergraduate students in humanities, social sciences or computer science, and were all native speakers of Swedish.

The 64 subjects in the conclusive study were assigned three out of 16 configurations of the game. The game and its configurations are described below. Each subject took part in three trials, thus each of the 16 configurations was presented to 12 different subjects, which resulted in a total of 192 trials in the conclusive study. The configurations and their order of presentation were equally and randomly assigned across random subjects with respect to gender. The total time to perform the experiment varied between different subjects, ranging between 15 to 30 minutes.

#### *Material*

A simple computer game was created in Macromedia FX. The task of the game is to steer a vessel within a fixed space. The game is completed through collecting four crosses scattered across the space and avoiding the present obstacles. The 2-dimensional game space is viewed from above and has five essential parts: the *vessel*, *obstacles*, *crosses*, and the *border of the game space* and a *counter* of the number of lives/attempts left. The vessel is controlled by three of the arrow keys on the keyboard: "arrow up", "arrow left" and "arrow right". "Arrow up" accelerates the vessel in the forward direction. "Arrow left" steers the vessel to the left with respect to the forward direction. "Arrow right" steers the vessel to the right with respect to the forward direction. Obstacles are stationary and located in four groups across the game space. There are four red crosses, which are collected by moving the vessel over them. The crosses are stationary and visible one at a time. When a cross is collected it disappears, and the next cross becomes visible until all the crosses are collected. Each cross appears in the same position every time. The game is played against a grey background (RGB: #CCCCCC) and a white border frames the area where the vessel may move from the surroundings. Each trial offers five attempts to collect the crosses, and the number of attempts left is shown on a counter located in the upper left part of the screen (within the area of movement). When failing one attempt, the crosses that already have been collected are kept and the playing continues from the starting point in the lower left part of the game space. Details of the game are shown in Figure 4 and 5. The size of the vessel in comparison to the game space is approximately 1:100.

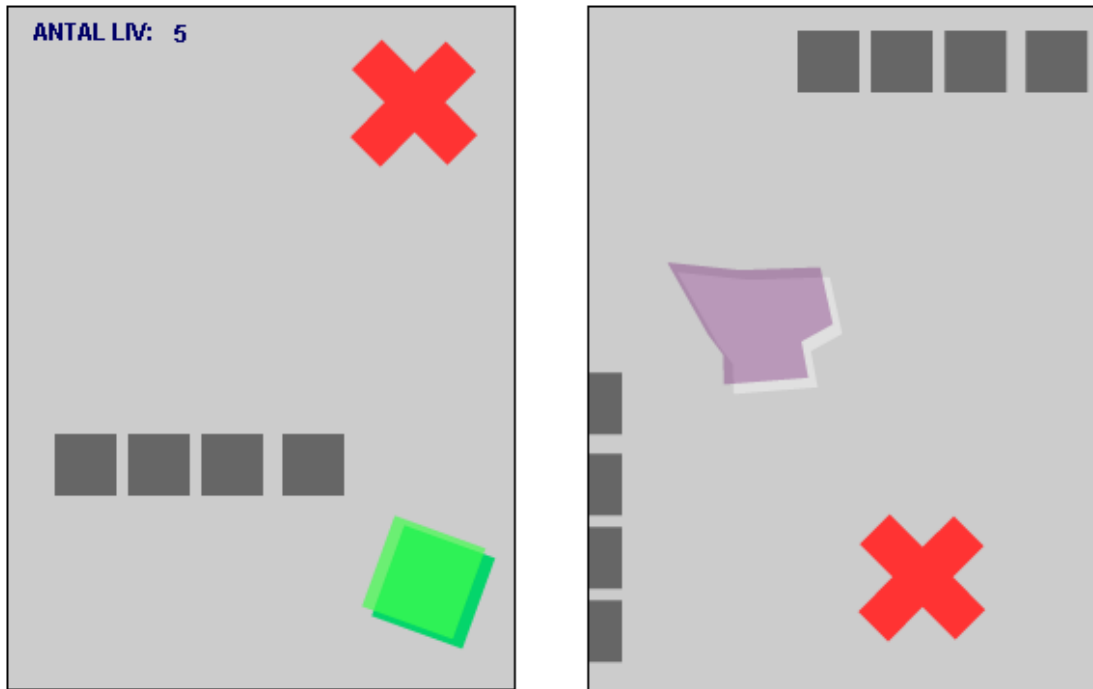


Figure 4 (left) and 5 (right): Details from the game. A cross appears in the upper right corner in the left picture. The vessel, with the appearance of a **square** is in the lower right of the left picture. The vessel has the appearance of an **amoeba** in the right picture where it is positioned in the middle. Obstacles are the squares visible on both pictures, and at the upper left corner of the left picture is the counter of the number of attempts left.

Four properties of the game systematically vary in two ways, which results in a  $2 \times 2 \times 2 \times 2$  factorial between-groups design and 16 different configurations of the game. The properties of variation are: the *visual appearance* of the vessel, *conditions of operation* of the vessel, *capriciousness in manoeuvring* the vessel and the *fragility* of the vessel. Once again the intent is not primarily to explore what specific properties elicits a characteristic to be ascribed to a program, but to elaborate the method able to account for these kinds of relationships. These configurations may to some extent be interpreted as including a variation of more than one quality within a configuration. This is deliberate and intends to make the configurations as different as possible with respect to the given program to elicit well separated impressions of the configurations. The later three properties of variation are not directly visible but are noticeable when manoeuvring, hence comprising the behavioural appearance of the vessel. The set of *visual appearances* are: a twinkling green **square** (RGB: #04D56B), with a thrust sound starting every time it's accelerated; a twinkling mauve **amoeba** (RGB: #A382A3), that changes shape<sup>1</sup>. The set of *conditions of operation* are: a **fast and agile** condition, with fast acceleration, high agility and maximal speed and with a slow decay in speed when not accelerating; a **slow and immobile** condition, with slower acceleration, less agility,

<sup>1</sup> The changing shape of the amoeba may be interpreted as behaviour. Using the dichotomy between visual and behavioural appearance as defined within this thesis, this appearance occurs without respect to the user's input to the process of interaction and is hence considered as merely visual in the description of the configurational properties.

low maximal speed and with a fast decay when not accelerating. The set of *capriciousness's in manoeuvring* the vessel are: **without disturbance** in the steering; **with disturbance** in steering that consists of a continuous random deflection from the forward direction. The set of *fragilities* of the vessel are: **no inherent fragility**; **with inherent fragility**, that make the vessel explode when accelerated or manoeuvred too violently, according to a prefixed ratio. The configurational properties were labelled according to a proper letter, consequently the label is S for the property **square**, A for **amoeba**, F for **fast and agile**, I for **slow and immobile**, W for **without disturbance**, D for **with disturbance**, N for **no inherent fragility** and Y for **with inherent fragility**. The configurations that are combined out of these configurational properties are labelled through a combination of the properties they possess, according to Table 7. These labels, descriptions and characterisations presented within the text are not shared with the subjects, but are used as a point of reference for own usage and presentation. It is possible that participants would not agree on some of the descriptions and characterisations of the different configurations.

Table 7: Labels for the four configurational properties and the combinations in 16 configurations. Abbreviations: S for **square** and A for **amoeba**; F for **fast and agile** and I for **slow and immobile**; W for **without disturbance** and D for **with disturbance**; N for **no inherent fragility** and Y for **with inherent fragility**.

	No inherent fragility				With inherent fragility			
	Without disturbance		With disturbance		Without disturbance		With disturbance	
	Fast and agile	Slow and immobile	Fast and agile	Slow and immobile	Fast and agile	Slow and immobile	Fast and agile	Slow and immobile
<b>Square</b>	<i>SFWN</i>	<i>SIWN</i>	<i>SFDN</i>	<i>SIDN</i>	<i>SFWY</i>	<i>SIWY</i>	<i>SFDY</i>	<i>SIDY</i>
<b>Amoeba</b>	<i>AFWN</i>	<i>AIWN</i>	<i>AFDN</i>	<i>AIDN</i>	<i>AFWY</i>	<i>AIWY</i>	<i>AFDY</i>	<i>AIDY</i>

The game was dimensioned 800x685 pixels and was displayed on a notebook with a 15" screen with a resolution of 1024x768 pixels. The game was shown as a Flash movie in Macromedia Flash through the command "Test Movie".

#### *Procedure*

The experiments were conducted in a secluded room at the Department of Computing Science at the University of Umeå. As introduction, subjects were reviewed about the procedure of the experiment and told that they were about to play three computer games. The subjects were verbally instructed about how to play the game and its controls through the arrow keys. This instruction employed a manuscript written in advance that describe how the game is completed through collecting four crosses within five attempts and avoiding the present obstacles. Subjects were told that these instructions were the same for all three games, and also to ask eventual questions before starting the trials. They were also informed they at any time could abort the experiment and that they were expected to answer a questionnaire upon each game. The experiment continued with the first game configuration that had been assigned to the subject. After the first trial,

subjects were given the questionnaire and read the written instructions on how to answer the questionnaire. These instructions describe the study, and how to answer the questionnaires by selecting the descriptive concepts that best correspond to their impression of the program on each pair. After this, subjects completed the first questionnaire, which took approximately 5 minutes. This process was repeated for the second and third trials. Subjects first played the assigned game configuration and then they completed the questionnaire on each configuration respectively. Concurrently with the experiment the number of crosses collected for each trial was registered.

### Results

The mean number of crosses collected in each configuration appears in Figure 5. This figure reveals that the least crosses are collected in the configurations **with disturbance** and **with inherent fragility**, i.e. *SIDY*, *AIDY*, *SFDY* and *AFDY* where the mean value is 0.00 or almost 0. On the other hand, the configurations *SIWN*, *AIWN*, *SFWN*, *SIDN*, *AIDN* and *SFWY* all have a mean value well above 3, when the maximal value is 4.00.

Totally four pairs in three questionnaires were not filled in by the subjects. A single missing data point results in that all answers in the whole questionnaire have to be excluded from the FA. As one single missing data point on a item alone only constitute approximately 0.5% of the total amount of answers to that item, it was decided to approximate these missing points with the alternative *Neither*.

Consequently, all 192 questionnaires with 39 dependent variables that correspond to each pair respectively were used for further analysis. Subject's ratings on the dependent variables were ranked between -3 and +3 in seven discrete steps, according to the same system used in the first study. Judging a pair by selecting the descriptive concept *Extremely* for the negatively valued word resulted in that the corresponding variable was ranked as -3. Contrary, judging *Extremely* for the positively valued word the corresponding variable was ranked as +3. A judgment of *Greatly* was ranked as -2 or +2, *Slightly* as -1 or +1 and *Neither* as 0. The normal distribution of the sample was derived out of the ratings on the dependent variables, with respect to its skewness and kurtosis. None of the variables did pass the Kolmogorov-Smirnov test for normality, using the

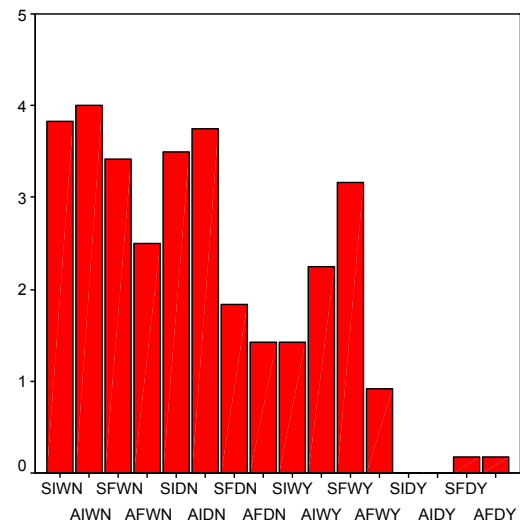


Figure 5: Mean number of crosses collected in each configuration. Abbreviations: S for **square** and A for **amoeba**; F for **fast and agile** and I for **slow and immobile**; W for **without disturbance** and D for **with disturbance**; N for **no inherent fragility** and Y for **with inherent fragility**. The maximal amount of crosses is 4. N=12 within each configuration.

conventional significance level of .01. A PCA extraction with Varimax rotation was hereafter conducted including all variables for all configurations. The extraction resulted in six components with eigenvalues over 1.0 that account for over 80 % of the total variance in the solution. The six components converged in 8 iterations with Varimax rotation are presented in Table 8.

The extracted component structure shows that 18 of the variables have primary loadings on the first component. Loadings on this component range between .90 and .58. Examples of pairs coinciding with this first component are RELIABLE–UNRELIABLE, CAREFUL–DISORDERED and SECURE–INSECURE. The second extracted component includes eight pairs of primary loadings between .90 and .46. Pairs loading on the second component include PLAYFUL–PRUDENT and VIGOROUS–COMPLIANT. Loadings for primary variables on the third component are ranging between .75 and .60. Pairs loading on this component are for example EFFECTIVE–IDLE and ACTIVE–PASSIVE. Loadings on component four range between .78 and .52, and include three pairs; among these pairs is SENSIBLE–SUPERFICIAL. Loadings on the fifth component range between .63 and .56, and also include three pairs with primary loadings. Among these are EASED–TENSE. Primary loadings on the sixth component come from two variables, INDULGENT–DOMINANT and OUTGOING–RESERVED with component loadings .58 and .42 respectively. Loadings further than the primary loadings are in general much lower than the primary loadings on all of first three extracted components. The difference between the loadings on the primary component and the other components are much less legible for the latter three components.

All pairs belonging to the original category *ACTIVITY* had their primary loadings on the same extracted component (the third), though once again with the exception of the pair VITAL–RESIGNED which has its primary loading on the second component. The primary loadings on the second component mainly come from pairs originating in the category *EXTRAVERSION* (six out of eight pairs) in the Big-Five inventory. The fourth component consists of three pairs of the original *INTELLIGENCE* category. The first component with 18 variables of primary loading, include constituting pairs of words that come from all of the original Big-Five categories and the *ADDITIONAL* category. Of its constituting variables, five pairs originate in the *CONSCIENTIOUSNESS* category and four pairs originate in the *AGREEABLENESS* category of the Big-Five inventory.

Representative words were selected as descriptors of each extracted component for further analysis. This was once again done according to their loadings on the component and due to their appropriateness as descriptors. The first component was hence classified as “Reliability”. The second component was classified as “Playfulness”, the third component as “Activity”, the fourth component as “Intelligence”, the fifth component as “Ease” and the sixth component as “Indulgence”. These component classifications are depicted in Table 9.

Table 8: Varimax-rotated component loadings derived in 8 iterations through PCA. Values equal to or larger than  $|\cdot 32|$  are listed in boldface type. Category of origin and index in the questionnaire are listed together with the name of the pair. Original category name abbreviations: *AGR* for *AGREEABLENESS*; *CON* for *CONSCIENTIOUSNESS*; *ACT* for *ACTIVITY*; *EXT* for *EXTRAVERSION*; *INT* for *INTELLIGENCE*; *EMO* for *EMOTIONAL STABILITY*; *ADD* for *ADDITIONAL*.

Pair	Category	Index	Component					
			1	2	3	4	5	6
RELIABLE – UNRELIABLE	<i>AGR</i>	20	<b>.90</b>	-.04	-.03	.06	.03	.08
HONEST – ERRATIC	<i>CON</i>	21	<b>.85</b>	-.06	-.02	-.04	.06	.14
CAREFUL – DISORDERED	<i>CON</i>	15	<b>.83</b>	-.15	-.05	.13	.02	-.23
SAFE – UNSAFE	<i>ADD</i>	34	<b>.82</b>	-.07	.05	.07	.19	.04
RELIEVED – TROUBLESOME	<i>EMO</i>	22	<b>.80</b>	.09	.04	-.04	.11	.11
ACCOMMODATING – DEMANDING	<i>AGR</i>	8	<b>.79</b>	-.11	-.13	-.14	.15	.06
CALM – IRRITABLE	<i>EMO</i>	10	<b>.79</b>	-.01	-.22	.16	.01	.22
SECURE – INSECURE	<i>ADD</i>	39	<b>.78</b>	-.07	.06	.19	.02	-.05
CONSIDERATE – NONCHALANT	<i>CON</i>	9	<b>.77</b>	.03	-.05	.15	.07	.20
OBLIGING – DISOBLIGING	<i>AGR</i>	14	<b>.76</b>	.10	-.02	.10	.26	.29
STEADY – SLOPPY	<i>CON</i>	3	<b>.75</b>	-.16	-.01	.07	.14	-.10
PLEASANT – UNPLEASANT	<i>AGR</i>	26	<b>.71</b>	.16	.07	.14	.24	.30
HARMONIOUS – CHAOTIC	<i>ADD</i>	33	<b>.70</b>	<b>-.33</b>	.05	.04	.26	.11
WISE – NAIVE	<i>INT</i>	5	<b>.67</b>	-.04	.15	<b>.37</b>	.10	.13
DETERMINED – UNCERTAIN	<i>EXT</i>	13	<b>.65</b>	-.04	.22	.18	-.05	-.29
REFINED – DULL	<i>INT</i>	11	<b>.63</b>	.05	.35	<b>.43</b>	.07	.02
CAUTIOUS – NEGLIGENT	<i>CON</i>	27	<b>.63</b>	-.30	.05	.17	.11	.08
RELAXED – NERVOUS	<i>EMO</i>	4	<b>.58</b>	-.08	-.20	.05	<b>.49</b>	.01
PLAYFUL – PRUDENT	<i>EXT</i>	35	-.04	<b>.90</b>	.14	.09	.12	.01
VIGOROUS – COMPLIANT	<i>EXT</i>	36	-.04	<b>.90</b>	.14	.09	.12	.01
ENTHUSIASTIC – BASHFUL	<i>EXT</i>	31	-.04	<b>.76</b>	.26	-.01	-.26	-.03
LIVELY – RESTRAINED	<i>ADD</i>	37	-.25	<b>.69</b>	<b>.34</b>	-.09	-.08	-.05
BOLD – UNOBTRUSIVE	<i>EXT</i>	7	-.15	<b>.64</b>	<b>.45</b>	.01	-.01	-.05
ELATED – RESTRAINED	<i>EXT</i>	25	-.15	<b>.60</b>	.19	.09	.10	.09
VITAL – RESIGNED	<i>ACT</i>	24	-.02	<b>.58</b>	<b>.41</b>	.26	.10	.27
MERRY – INTROVERTED	<i>EXT</i>	19	.05	<b>.46</b>	.26	.26	-.05	<b>.46</b>
EFFECTIVE – IDLE	<i>ACT</i>	12	.16	.28	<b>.75</b>	.01	-.06	.28
ACTIVE – PASSIVE	<i>ACT</i>	6	-.07	<b>.43</b>	<b>.71</b>	.08	.01	.00
PERFORMING – LACKING INITIATIVE	<i>ACT</i>	38	-.04	<b>.42</b>	<b>.71</b>	.01	-.05	-.04
ENTERPRISING – UNENTERPRISING	<i>ACT</i>	18	-.12	.28	<b>.67</b>	.14	.14	-.05
ENERGETIC – APATHETIC	<i>ACT</i>	30	.15	<b>.46</b>	<b>.60</b>	.16	.01	.02
SENSIBLE – SUPERFICIAL	<i>INT</i>	23	.16	.16	.03	<b>.78</b>	.17	.07
INTELLIGENT – UNINTELLIGENT	<i>INT</i>	17	<b>.52</b>	.08	<b>.37</b>	<b>.53</b>	-.03	.02
COMPETENT – IGNORANT	<i>INT</i>	29	<b>.51</b>	.04	<b>.41</b>	<b>.52</b>	-.06	-.02
EASED – TENSE	<i>EMO</i>	28	<b>.47</b>	-.04	.02	.16	<b>.63</b>	-.11
WARM – COLD	<i>AGR</i>	2	.18	.26	.21	.12	<b>.59</b>	.22
SERENE – IMPATIENT	<i>EMO</i>	16	<b>.51</b>	<b>-.32</b>	-.19	-.10	<b>.56</b>	.09
INDULGENT – DOMINANT	<i>AGR</i>	32	<b>.48</b>	-.25	-.08	-.10	.21	<b>.58</b>
EXTROVERTED – INHIBITED	<i>EXT</i>	1	.27	<b>.35</b>	.25	.22	.08	<b>.42</b>

Table 9: Showing classifications for components extracted in PCA with Pairs represented in each category.

Component classification	Representing pairs
“Reliability”	RELIABLE – UNRELIABLE, HONEST – ERRATIC, CAREFUL – DISORDERED, SAFE – UNSAFE, RELIEVED – TROUBLESOME, ACCOMMODATING – DEMANDING, CALM – IRRITABLE, SECURE – INSECURE, CONSIDERATE – NONCHALANT, OBLIGING – DISOBLIGING, STEADY – SLOPPY, PLEASANT – UNPLEASANT, HARMONIOUS – CHAOTIC, WISE – NAIVE, DETERMINED – UNCERTAIN, REFINED – DULL, CAUTIOUS – NEGLIGENT, RELAXED – NERVOUS.
“Playfulness”	PLAYFUL – PRUDENT, VIGOROUS – COMPLIANT, ENTHUSIASTIC – BASHFUL, LIVELY – RESTRAINED, BOLD – UNOBTRUSIVE, ELATED – RESTRAINED, VITAL – RESIGNED, MERRY – INTROVERTED.
“Activity”	EFFECTIVE – IDLE, ACTIVE – PASSIVE, ENTERPRISING – UNENTERPRISING, PERFORMING – LACKING INITIATIVE, ENERGETIC – APATHETIC.
“Intelligence”	SENSIBLE – SUPERFICIAL, INTELLIGENT – UNINTELLIGENT, COMPETENT – IGNORANT.
“Ease”	EASED – TENSE, WARM – COLD, SERENE – IMPATIENT.
“Indulgence”	INDULGENT – DOMINANT, EXTROVERTED – INHIBITED.

Similar to the procedure of the initial experiment, the rankings on the components between the different configurations were estimated through the regression approach, to gain a counterpart to the classified components as a single item of direct evaluation. These estimated component scores were analysed through a linear regression model to explore differences in subject’s impressions of the program according to its configurations. An analysis of variance between groups (ANOVA) show significant differences between the configurations on the impression of the component “Reliability” ( $p=.000$ ) and “Intelligence” ( $p=.009$ ). No other significant differences are found between the configurations. These results are presented in Table 10.

Table 10: ANOVA showing differences in the ascription of characteristics to the program, according to the configurations (N=192). Total SS for each component respectively is 191.0. Significant differences ( $p < 0.05$ ) are listed in boldface type.

	“Reliability”	“Playfulness”	“Activity”	“Intelligence”	“Ease”	“Indulgence”
F	12.85	1.20	.99	2.16	1.24	.80
Sig.	<b>.000</b>	.275	.466	<b>.009</b>	.247	.682

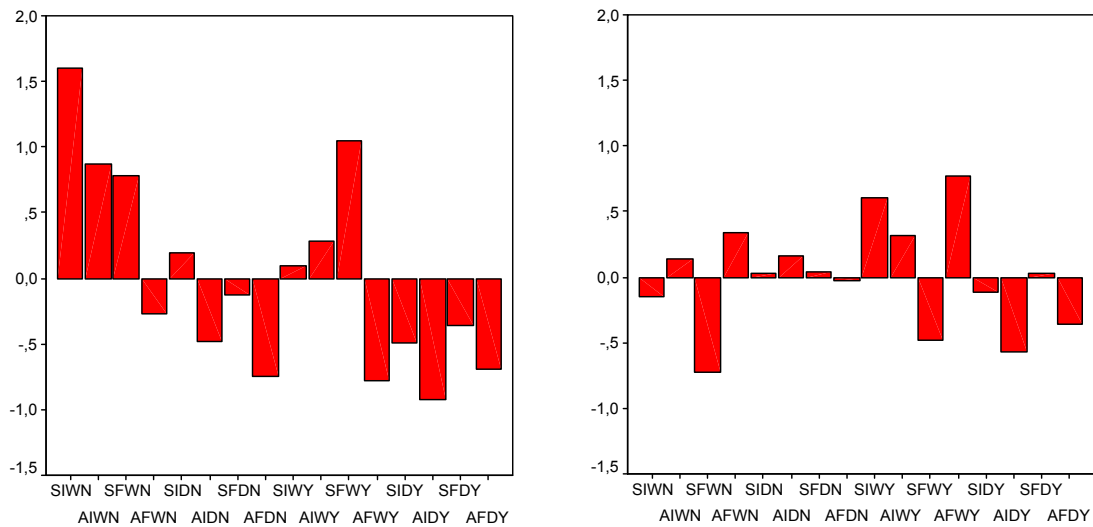


Figure 6 (left) and 7 (right): Estimated mean scores for the components “Reliability” (to the left) and “Intelligence” (to the right) due to configuration. Abbreviations: S for **square** and A for **amoeba**; F for **fast and agile** and I for **slow and immobile**; W for **without disturbance** and D for **with disturbance**; N for **no inherent fragility** and Y for **with inherent fragility**. Maximal score = 3, minimal score = -3.

The estimated mean scores for the components “Reliability” and “Intelligence” are presented in Table 11 and depicted in Figures 6 and 7. As shown by Table 11 and Figure 6, mean scores for the component “Reliability” are highest for the configurations *SIWN* (Mean = 1.597), *SFWY* (Mean = 1.044), *AIWN* (Mean = .875) and *SFWN* (Mean = .776) whereas mean scores are lowest for the configurations *AIDY* (Mean = -.925), *AFWY* (Mean = -.783), *AFDN* (Mean = -.742) and *AFDY* (Mean = -.689). The total SS is 191.0, the maximal mean value is +3 and the minimal mean value is -3. The data show the configurations **without disturbance** in general are evaluated as more positive on the component “Reliability” in comparison with the configurations **with disturbance**. Only one configuration **with disturbance**, *SIDN* have a mean score higher than zero (.192) on the component “Reliability”. There are several statistical significances on the  $p < .05$  level between the mean scores for these configurations that score “high” and those that score “less high” on the component “Reliability”. For example, there is a statistical significance on the  $p < .05$  level between the mean scores for the configuration *SIWN* and the configurations *AFWN*, *AIDN*, *SFDN*, *AFDN*, *SIWY*, *AFWY*, *SIDY*, *AIDY*, *SFDY* and *AFDY* respectively. On the component “Intelligence” the most differential mean scores are between the configurations *SFWN* (Mean = -.730) and *AFWY* (Mean = .771). The difference between these mean scores for these configurations on the component “Intelligence” is not significant ( $p = .136$ ). ANOVA on the differences for the mean scores on the four other components show no statistical significances between any configurations, and these results are hence not included.

Table 11: Estimated mean scores and standard deviations for the eight configurations on the components “Reliability” and “Intelligence”. Positively valued scores are marked in boldface type. Maximally positive score is +3; maximally negative score is -3. Abbreviations: S for **square** and A for **amoeba**; F for **fast and agile** and I for **slow and immobile**; W for **without disturbance** and D for **with disturbance**; N for **no inherent fragility** and Y for **with inherent fragility**.

	<i>SIWN</i>	<i>AIWN</i>	<i>SFWN</i>	<i>AFWN</i>	<i>SIDN</i>	<i>AIDN</i>	<i>SFDN</i>	<i>AFDN</i>
“Reliability” Mean	<b>1.597</b>	<b>.875</b>	<b>.776</b>	-.267	<b>.192</b>	-.479	-.129	-.742
“Reliability” Std. Dev.	.798	.683	1.032	.566	.993	.708	.771	.279
“Intelligence” Mean	-.149	<b>.143</b>	-.730	<b>.337</b>	<b>.031</b>	<b>.158</b>	<b>.045</b>	-.025
“Intelligence” Std. Dev.	1.051	1.100	.861	.649	.994	1.132	.754	1.204
	<i>SIWY</i>	<i>AIWY</i>	<i>SFWY</i>	<i>AFWY</i>	<i>SIDY</i>	<i>AIDY</i>	<i>SFDY</i>	<i>AFDY</i>
“Reliability” Mean	<b>.092</b>	<b>.288</b>	<b>1.044</b>	-.783	-.494	-.925	-.356	-.689
“Reliability” Std. Dev.	.887	.840	.823	.675	.564	.518	.428	.497
“Intelligence” Mean	<b>.605</b>	<b>.315</b>	-.485	<b>.771</b>	-.118	-.567	<b>.030</b>	-.361
“Intelligence” Std. Dev.	.913	.624	1.089	1.090	1.030	.998	.842	.730

A multivariate general linear model provides regression analysis and analysis of variance that show differences in the impressions according to the set of the four configurational properties **visual appearance**, **conditions of operation**, **capriciousness in manoeuvring** and **fragility** and also effects of interaction between these properties. This renders the possibility to associate specific configurational properties to specific impressions of the stimuli program among the subjects. The results from this analysis indicate a significant difference in the impression of the stimuli program between the set of configurations of **visual appearance** on the components “Reliability” (p=.000) and “Ease” (p=.006). The configurationally property **square** is considered as more “Reliable” in comparison with the property **amoeba**. This latter property is however considered as more “at ease” in comparison with the property **square**. The analysis also shows a significant difference in the impression of the stimuli program between the **conditions of operation** on the components “Reliability” (p=.047), “Playfulness” (p=.002) and “Activity” (p=.001). The **slow and immobile** property appear to be considered as more “Reliable” in comparison to the **fast and agile** property whereas the latter is considered by the subjects to be both

more “Playful” and “Active” in comparison to the former. Furthermore, there is a significant difference according to the component “Reliability” ( $p=.000$ ) with respect to the property *capriciousness in manoeuvring*, where the configurations **without disturbance** is considered as more “reliable” than the configurations **with disturbance** in the manoeuvring. Finally, the analysis also show a significant difference in the impression of the stimuli program between the set of properties of **no inherent fragility** and **with inherent fragility** on the components “Reliability” ( $p=.001$ ) and “Indulgence” ( $p=.035$ ). The configurations with **no inherent fragility** are considered as more “reliable” and “indulgent” in comparison to the configurations **with inherent fragility**. These findings are presented in Table 12.

Table 12: General linear model analysis showing differences in the ascription of characteristics to the program, according to it’s configurationally properties (N=192). Significant differences ( $p<.05$ ) are listed in boldface type.

Components	Property: <i>Visual appearance</i> (df=1)		Property: <i>Conditions of Operation</i> (df=1)		Property: <i>Capriciousness in manoeuvring</i> (df=1)		Property: <i>Fragility of the vessel</i> (df=1)	
	F	Sig.	F	Sig.	F	Sig.	F	Sig.
“Reliability”	25.01	<b>.000</b>	4.00	<b>.047</b>	49.30	<b>.000</b>	10.47	<b>.001</b>
“Playfulness”	.19	.660	10.18	<b>.002</b>	.03	.853	.00	.956
“Activity”	.17	.681	10.68	<b>.001</b>	.54	.462	.18	.674
“Intelligence”	1.79	.182	.52	.471	1.96	.163	.11	.743
“Ease”	7.72	<b>.006</b>	.52	.474	.25	.617	1.49	.223
“Indulgence”	.69	.407	.05	.828	.84	.360	4.53	<b>.035</b>

Several effects of interaction are found between the properties in the analyzed material, which are briefly presented here being considered as out of the general scope of investigation. There is a significant interaction effect with respect to the component “Intelligence” between the set of configurationally properties *visual appearance* and *conditions of operation* ( $p=.050$ ), *visual appearance* and *capriciousness in manoeuvring* ( $p=.006$ ), *capriciousness in manoeuvring* and *fragility* ( $p=.011$ ) and between the set of properties of *visual appearance*, *conditions of operation* and *capriciousness in manoeuvring* ( $p=.047$ ). There is also a significant interaction effect with respect to the component “Reliability” between the set of configurationally properties *visual appearance* and *conditions of operation* ( $p=.009$ ), *conditions of operation* and *capriciousness in manoeuvring* ( $p=.026$ ), *conditions of operation* and *fragility* ( $p=.001$ ), *visual appearance*, *conditions of operation* and *capriciousness in manoeuvring* ( $p=.003$ ) and between the set of properties *visual appearance*, *conditions of operation* and *fragility* ( $p=.047$ ). Finally, there is also a significant interaction effect with respect to the component “Ease” between the set of the configurational properties *conditions of operation* and *fragility* ( $p=.022$ ).

### *Discussion*

Results from the second experiment indicate a notable difference in the achievement of the subjects between the different configurations of the stimuli program. Even though the performance of the participants was not directly in the focus of the investigation, it is interesting but not surprising that subjects in general achieve much better in configurations **without disturbance** and **without fragility** in comparison to configurations with these properties. Moreover, it is apparently not impossible to manage the vessel in some configurations **with disturbance** or **with inherent fragility**. The mean scores of collected crosses is among the highest for the configurations *SIDN*, *AIDN* and *SFWY* – configurations either **with disturbance** (abbreviated D) or **with inherent fragility** (abbreviated Y). The achievement rather appears to be a matter of the combination of properties. The mean scores for the number of crosses collected within each configuration may serve as an indication of how a formal measurement of performance is connected to the impression of characteristic attributes of the stimuli program. As indicated by Figure 5 and 6, there seems to be a slight correspondence with respect to the achievement in collecting crosses and the impression of “Reliability”. In general, there is a correspondence in the achievement (i.e. with respect to collecting crosses) on the configurations with high ratings on “Reliability”.

Unfortunately, the dependent variables derived in the experiment show a variance that diverges from a sample of normal distribution. The statistical power of the FA and the mean component scores estimated out of the analysis is hence a source of concern. Presenting a material for experimental evaluation in which the dependent variables that are measured show a normal distribution may be difficult in a setting of subjective measurement, as shown by the present experiment. There is however still reasons to investigate the underlying component structure through a FA – above all as a means to explore implications for a general inventory for assessment of characteristics in screen-based computer artifacts.

These components derived in the second experiment show some interesting findings concerning the correlations among the dependent variables with respect to the implications from the first experiment. First of all, the correlation among the words describing “Activity” that was found in the first experiment is reproduced in this experiment, which was conducted within a different domain. These words exclusively load on the third extracted component with the exception of the pair VITAL–RESIGNED as depicted in Table 8. In the first experiment, the component “Activity” also included words originating in the Big-Five category *AGREEABLENESS*, such as WARM–COLD. This was not the case for the second experiment. Here, this specific pair had its primary loading on the fifth extracted component together with two pairs originating in the category *EMOTIONAL STABILITY*. In the initial experiment, the extracted component “Relaxation” included five pairs originating in the *EMOTIONAL STABILITY* category, of which three now have their primary loadings on the component “Reliability” of the second experiment. Of these, only the pair RELAXED–NERVOUS show a fair secondary loading (.49) on the component that is classified as “Ease” in the second experiment. The extracted component “Intelligence” is preserved between the two experiments, but whereas there are six pairs with primary loadings on this component in the first

experiment there are three pairs in the second. The pairs with the lowest primary loadings on the component “Competence” in the first experiment have their primary loadings on the component “Reliability” in the second experiment. Although, two of these pairs still have relatively high secondary loadings on the “Intelligence” component (i.e. WISE–NAIVE and REFINED–DULL). The component “Playfulness” has its highest primary loadings from four of the new pairs (PLAYFUL–PRUDENT, VIGOROUS–COMPLIANT, ENTHUSIASTIC–BASHFUL and LIVELY–RESTRAINED). The rest of its constituting pairs were not coinciding in the extracted components of the first experiment. Consequently, it seems that the addition of the new pairs may have contributed to catch a dimension of the semantic space that is disregarded within the first experiment, which is expressed by that pairs that were uncorrelated the first experiment now correlates on the “Playfulness” component.

Furthermore, the first component extracted in the PCA includes 18 pairs of words. Among these 18 pairs are the 5 pairs that were extracted for this component in the first experiment. This component was classified as “Reliability” in both experiments and its constituting pairs of words comes from all of the original Big-Five categories, with a predominance of pairs originating from the categories *CONSCIENTIOUSNESS* (five pairs) and *AGREEABLENESS* (four pairs), but also from the new *ADDITIONAL* category. However, as only three of the pairs with primary loadings on this component are from this “Additional” category, there are consequently ten pairs that have switched its component of primary loading. Another interesting observation is that many of the pairs that in the second experiment load on the “Reliability” component in the first experiment have loadings greater than  $|\text{.32}|$  on the “Reliability” component of that experiment. As pointed out by Goldberg (1992), descriptive words in the Big-Five inventory tend to share some meaning with more than one factor. A change in loadings between components through the two experiments may hence find explanation in that the meaning of descriptive words is subject to change between different domains of evaluation. There is also a source of insecurity in the translation of words used in the questionnaire from English to Swedish, which may change their semantic content in an uncontrolled fashion. Concluding, there seems to be an overlap in the loadings on the extracted components between the two experiments with regard to some components, as with “Activity”, but this is not a general tendency. For most components, the pattern is that they are preserved, but that some or many of the pairs loading on the component changes, as with the components “Reliability” and “Intelligence”. This indicates that the domain of personality and the present screen-based computer programs are not completely equal objects of evaluation.

Analysing the estimated component scores with ANOVA shows there is a significant difference in subjects’ ascription of “Reliability” and “Intelligence” due to configuration of the stimuli program. The statistical power in these analyses is satisfactory as the F statistic is robust even to unequal variances when sample sizes are equal. There are several statistical significances between the different configurations with respect to the component “Reliability”. Most notable is the difference in the impression of the configurations *SIWN* and *AIDY*, which have the most positive and least positive mean values on this component. However, there are several other statistically significant

differences between the configurations on the component “Reliability”. Apparently, the configurational property **with disturbance** seems to influence the impression of “Reliability” most negatively, as it is mostly these that are considered as least “reliable”. In addition, the configurational property **with inherent fragility** as a general tendency only seem to influence a particularly negative impression of “Reliability” together with configurations with the property **with disturbance**. Concerning the component “Intelligence”, the difference in evaluation between most differential mean scores is not statistically significant ( $p=.136$ ), hence the difference between the configurations on this component apparently is less explicit but nevertheless statistically significant on a level of general comparison. There are however, as discussed ahead, several statistically significant effects of interaction between the set of configurational properties on the component “Intelligence”.

The general linear model analysis shows a statistically significant difference between each set of configurational properties on the extracted component “Reliability”. The impression of “Reliability” hence seems to describe the stimuli program concerning its controllability and predictability as the properties **square, slow and immobile, without disturbance** and **without inherent fragility** all influence a more positive impression of “Reliability” in the stimuli program, in comparison to each of their opposite properties respectively. Moreover, there are other statistically significant differences between the set of configurational properties, perhaps most interestingly in the impression of “Playfulness” between the set of *conditions of operation*. The property **fast and agile** is considered as more “Playful” and “Active” in comparison to the property **slow and immobile**. A possible interpretation of this finding is that the property **fast and agile** evokes this kind of impression due to the fact that it provokes more occurrences and action as a result of its speed and less strict controllability. A difference in the impression of the component “Ease” is statistically significant for the configurational property *visual appearance*. This finding might be explained as due to the fact that the continuously changing shape of the **amoeba** makes its appearance more “eased” and “warm”. Finally, there is also a significant difference between the set of configurational properties of **fragility of the vessel** on the component “Indulgence”. The matter that the configurations with **no inherent fragility** are considered as more “indulgent” in comparison to the configurations **with inherent fragility** might be explained as due to the fact that their relative stability, not causing unpredictable explosions, influence this impression.

There are also several significant interaction effects between the configurational properties that primarily generates differences in the ascription of “Reliability” and “Intelligence” among the subjects. Without discussing these effects extensively, it may be concluded that it apparently is possible that configurational properties, which alone do not produce statistically significant differences in impression might do this in interaction. This is especially apparent with the difference in the impression of the component “Intelligence”. This matter indicates the importance to further investigate how the diverse characteristics ascribed to artifacts may together form a more coherent ascription of the whole character, which is not investigated in the present study.

### *Concluding discussion*

The two experiments conducted within the frame of this thesis seem to strengthen the assumptions of Janlert and Stolterman's (1997) approach on an ascription of character to computer artifacts. The result of the first experiment indicates that the participants in general agree to use descriptive concepts associated with the lexical descriptors used in the questionnaire to describe their impression of the stimuli program of these experiments. There is a concordance among the participants of the experiments in the evaluation of the different configurations on the contrasting semantic pairs in both experiments. This is expressed through the fact that it is possible to derive significant differences in the impressions of various components between specific configurations and sets of configurational properties. This piece of information makes it possible to relate the underlying relations between the variables in the subset to systematic differences between the impressions of different sets of configurational properties in the stimuli program. The results from these experiments hence show that it is possible to successfully find differences in the impression of and ascriptions of characteristics to screen based computer programs through the elaboration of its properties when evaluated through the semantic differential technique. Moreover, the result also shows the possibility to assess significant differences in impressions and ascriptions of characteristics between the configurations with respect to the diverse qualities of the stimuli programs. As indicated within this study, this may be possible not only concerning the functionality and visual appearance, but also with respect to the behavioural appearance of an artifact.

As a starting point for the present investigation a restricted range of contrasting semantic pairs were selected for evaluation through the semantic differential technique. In order to develop a general procedure for evaluation and assessment of characteristics a FA was used to investigate the underlying relations between these pairs. The components extracted in this analysis are constituted of coherent subsets with high internal correlation and which are relatively independent of other subsets. The rationale of this approach is to provide a more compact tool of evaluation that summarizes the most essential parts of the characteristics ascribed to an artifact. The questionnaire developed within this thesis should not be considered as the final tool for this procedure, but the components identified and the words used within the two experiments may serve as a foundation for future work. There are several issues of methodological and theoretical concern, for which improvements are to be discussed below. The greatest relevance of this presented questionnaire and its content is rather to point out a direction, and to serve as a valuable input and illustration of a general procedure to produce a tool for evaluation of the characteristics ascribed to a specific object of interest.

The supplementary category *ACTIVITY* has a stable reproduction between the two experiments conducted within this study. Causes for this result may be that this category has an underlying set of contrasting semantic pairs that are very similar in meaning, but it might also be that this the meaning of category is easy to identify, evaluate or experience in the setting whereas the opposite may be true for other categories. The need of further exploration of more descriptors, as a means to investigate a broader semantic space describing the characteristics of artefacts is apparently significant. This was expressed by the finding of the "Playfulness" component within the second experiment – the addition

of the new pairs seems to catch a dimension of the semantic space that is disregarded in the first experiment. Consequently, a general implication for further studies is to initially collect a much larger inventory of relevant words that go well beyond the Big-Five categories and which may describe several other aspects of the semantic space than these words does. A too small inventory of words in the questionnaire also runs the risk of producing rather homogeneous factors that provide low degree of reliability as pointed out by Saucier (1994).

The rationale of using contrasting semantic pairs as a basis for assessment is to better specify the item of evaluation for each scale, as the interpretation of a single word differs among people. There are several doubts and considerations to this approach, though. First of all, the participants may under some circumstances not experience the pairs matched as semantic contrasts as opposites, and hence it will not be appropriate to evaluate them as such. In some cases, for example, it would be more feasible to give a similar judgment on both words within the pair. Second, in some cases the evaluative dimensions between the contrasting words within the pair are clear and obvious, with respect to which word is regarded as positively loaded and which word is regarded as negatively loaded. However, this is not the case for all pairs within the present questionnaire. As the positive and negative poles are not self evident within some pairs in the questionnaire, this may indicate that the ratings on the different pairs are not compared and analysed along the similar scale of evaluation. Third, a study by Goldberg (1992) did indicate that unipolar variables provide a more univocal representation of the Big-Five factor structure than a bipolar scale did at a comparison. Consequently, these considerations reveal that the use of a unipolar scale may provide a more satisfactory format of evaluation. If desired, this may still provide a measurement on contrasting dimensions by including contrasting semantic words and matching their results after evaluation of the questionnaire. More extensive exploration of the meanings and valences of different words within different domains and among different people are also desirable. A further concern of the evaluative procedure is also to in greatest possible extent using words that have a semantic meaning that is not directly positively or negatively valued. This may avoid that people describe an artifact just according to their positive or negative feelings towards it when a positive or negative meaning is very clearly expressed by the words that are used.

If possible there are good reasons to change the experimental procedure so that two separate parts of evaluation are performed similar to the methodology of KE, rather than comparing the configurations on the basis of the estimated component scores. Such procedure would benefit by providing direct judgments of the most essential aspects of an artifacts appearance. Amendment of the procedure for assessment of characteristics may hence be attained by an initial investigation through FA, followed by a separate experiment for direct evaluation of the most relevant semantic dimensions and fitting descriptors extracted in the factorial analysis. In a more extensive approach involving more participants, it might also for example be possible to investigate differences in the extracted structure between the different configurations of evaluation. This was not carried within the present study due to a limited number of subjects within each configuration. It is not unlikely that acquaintance with the descriptive concepts makes the evaluation of the artifacts easier over time.

Although supported by a data material with a deviant normal distribution, which conveys limitations to the results from the FA, a conclusion from these two experiments is that the different categories within the Big-Five factor structure are not clearly reproduced within the context of screen based computer artifacts. One conceivable explanation that must be taken into consideration is that the translation of the words used from English to Swedish may influence their meaning and change their underlying structure. Hence this matter of translation might to some degree affect the difference in the underlying structure. There is however some cross-cultural studies conducted within the research on the Big-Five factor structure, which indicate that the Big-Five factor dimensions are preserved when investigated within different languages (cf. McAdams, 1992). For future research of the structure among characteristics, it is however desirable that the semantic descriptors are operationalised for an improved normality distribution of the sample, adding a more satisfactory validity to the extracted components. Steps in this direction may be possible through the introduction of a refined experimental procedure as discussed above.

In the view of Janlert and Stolterman (1997), the character of an artifact is an emergent unity constituting of a number of characteristics that it displays. Within this thesis, the importance and interaction between different characteristics in forming a unified impression of character is not explored as the thesis concerns an investigation of the assessment of characteristics ascribed to artifacts. Consequently, at this stage it is not possible to conclude how the overall impression may influence the relationship between an artifact and its user. In Janlert and Stolterman's notion, the consistency of character is an important implication for the design of computer artifacts. Such a concern is however in general disregarded within the empirical evaluation of computer systems. Applying the procedure developed within this thesis as a starting point, an understanding of the impression of the whole character may be elaborated through the relationships between the distinct components that account for the most important dimensions of the semantic space. As these components may be derived through a FA the continuation of the process is consequently a matter of studying the coherence of different characteristics, and how, more generally, different characteristics work together to produce a unified character. The studies of Asch (1946) and Asch and Zukier (1984) draw some interesting inferences in that the attribution of character to people may be a emergent rather than simply additive process, and they also imply that there is a process of discrimination between central and peripheral traits, that cognitive content and functional value may change due to environment and that the impression of traits is a dynamic process in that some traits may determine both the content and the function of other traits. The investigations on the structure of personality character conducted by Asch and Asch and Zukier may consequently be an interesting starting point for practical considerations of the methodological procedure for elaboration with the unified character ascribed to artifacts.

As previously recognized within KE, the current study indicates that it is possible to discriminate among the impressions of the configurations of the stimuli programs according to differences in visual appearance. The findings of the present study also indicate that participants may form different impressions due to different qualities in functionality and behavioural appearance. According to Janlert and Stolterman's (1997)

notion as introduced above, attribution of character to artifacts may also generate expectations that help to anticipate behaviour, generate explanations of behaviour, serve as a context for interpretation and also serve as a schema for normal functioning. Applying the procedure for assessment that is developed within this thesis for investigation within these different functional areas of character attribution is an important area of future research. As shown within the second experiment, the impression of “Intelligence”, “Reliability” and “Playfulness” in the stimuli program is influenced by its various properties. As an example of Janlert and Stolterman’s view, if a user understands an artifact as reliable, this will guide the expectations on its behaviour. For example, if something unexpected or unknown appears during interaction this may not worry the user. It is also conceivable that the behaviour of such an artifact will be explained according to this impression and ascription of characteristics, and then if the artifact behaves unexpectedly, this may indicate that something is wrong. On the other hand, if the impression of an artifact is that it is a playful one, this may generate a tolerance if its behaviour is unanticipated. How ambiguousness in the appearance of an artifact affects its users is both of practical concern and theoretical interest. Janlert and Stolterman also claim that an impression of characteristics may encompass this whole range of different qualities of an artifact and the overall character ascribed to an artifact may cover all of these diverse qualities in unity. Understanding the relationships between the impressions of different qualities of an artifact is another matter of interest for future elaboration of the approach. This may concern the consistency and transferral of characteristics knowledge across different segments of an artifact’s manifestation and appearance, as from visual appearance to behavioural appearance, and from behavioural appearance to strategy of use.

For further investigation on the notion of the character of things as put forward by Janlert and Stolterman (1997) there are several other important implications as well. One limitation of the two experiments conducted within this study is the fact that each participant interacts with the stimuli program during a restricted range of time. The time for each trial is not settled but differs between participants. Thus, this might result in a difference in the familiarity each participant develops with the object of assessment. There are also important contributions and benefits to the application and applicability of the methodology if implemented in combination with other accounts. Making a brief outline of such an effort, input from the activity theory approach may stress the importance of context of use and the developmental perspective are important means for a complete picture in understanding the interaction between humans and artifacts (cf. Nardi, 1996). A development of the skills of using an artifact may coincide with a change of the ascriptions and the understanding of its characteristics, which may be manifested in terms of for example expectations and explanations generated about it. There is also a valuable input to the understanding of the characteristics ascribed to computer artifacts that might be derived from both novices and experienced users in applied settings. The focus on actual practice as a means to get knowledge on social understandings within the ethnomethodological approach (cf. Garfinkel, 1967) is hence also of relevance. The methodological procedure presented within this thesis may hence work as a point of departure for such more general investigations. By establishing a general structure of the essential characteristics ascribed to an artifact with this procedure, direct studies and

observations of user practice may be supported and more knowledge about the circumstances of and prerequisites for ascribing characteristics to artifacts might be gained.

In the approach on the character of things as presented by Janlert and Stolterman (1997) they argue that the ascription of character is not solely a linguistic phenomenon. There is an interesting parallel to the discussion on the notion of metaphors and analogies within systems design, as for example put forward by Halsaz and Moran (1981) on this matter. According to Halsaz and Moran, even though analogies are a natural and spontaneous way for people to think about a complex system that is poorly understood they may act as barriers that prevents users from getting a more deep and effective understanding in the long run. If developing applications and tools through approach and highlighting their character, it is essential that the character acts as a means to facilitate the conceptual understanding of the artifact. A certificate of the characters impact on such an understanding may in the end be expressed through its influence on the course of interaction, which is not highlighted in the present study. To verify and investigate this matter is consequently a great challenge for the further consolidation of the approach of the character of things approach as presented by Janlert and Stolterman. This may specifically be an investigation of the differences in styles of interaction and use with respect to the characteristic traits ascribed to an artifact by its users. Such studies may be conducted both through studies in the field, as discussed above and within more strictly implemented laboratory experiments. In both cases, video recordings of the situation of use may support observations and be combined with interviews. Formal measurements may be received through registration of the course of interaction within the artifact in use. Also, there is much research to be done on the ascriptions of characteristics due to determinants of personality and social and cultural differences.

The current procedure of assessment may provide a general indication on a relationship between discrete properties of an artifact and its correspondence to an impression of characteristics. To employ a design methodology for computer artifacts in which the character concept works as a mental aid that helps the user at interaction, findings concerning the relationship between ascriptions of characteristics and corresponding styles of interaction are of great benefit. In compliance with this effort, connections between specific attributions of characteristics to specific features of an artifact's appearances may be possible to derive. Applying the findings of the current investigation to the practical design work may hence as an example suggest that an artifact, for which it is desirable to behave in a playful manner, may have a more dynamic and less strictly controllable mode of interaction, as derived from the difference between the set of ***conditions of operation*** properties in the second experiment. The concept of producing reliable applications is a widely acclaimed goal within systems design. The great impact of the component "Reliability" in the second experiment may in one sense support this effort, as the sense of "Reliability" appears to be very important characteristic of this stimuli program. Also, this finding may indicate that the concept "Reliability" is a more complex notion than generally assumed as it includes a broader meaning than which is commonly associated with reliability.

The present thesis is concerned with the investigation of characteristics ascribed to a set of configurations of simple screen-based computer programs. As introduced above, one of the underlying assumptions of the investigation is that these ascriptions are easily made even in settings where little information is displayed in the observed object. Screen-based computer programs may be seen as a very special form of computer artifacts: at the same time as they represent a high complexity and variation in their processes, the tools for interaction are commonly restricted to quite simple forms of input and output devices as monitors and keyboards and mice. For experimental concerns, investigating screen based computer artifacts is rewarding, as it is easy to produce different prototypes of various configurations to undertake controlled experiments. However, it is necessary to not let this relative simplicity with screen based computer artifacts alone guide the exploration of the theoretical framework of the character of things. Exploring character ascriptions in other domains of computer artifacts may hence give occasion to explore their connections to a broader and more direct range of interaction strategies on the one hand and technical functionalities, visual and behavioural appearances on the other. The application and output of this approach may also be much more useful within other areas of life, where complicated, technical artifacts are present, as with cars, control rooms for supervision of industrial processes or mobile devices, to give a few examples of such possible areas.

Concluding, the present thesis has presented an initial empirical exploration of the character of things theory as put forward by Janlert and Stolterman (1997). The thesis presents a general procedure of empirical measurement and evaluation of characteristic attributes in computer artifacts that may connect specific impressions of characteristics to specific design properties. There are however several limitations and possible refinements to the present procedure as discussed above, but the results of this thesis nevertheless indicate that further empirical testing, exploration, elaboration and development of the character of things approach may serve as an important complement and contribution to other approaches in both practical design work and theoretical frameworks for understanding human abilities and the human mind.

## References

- Asch, S. E. (1946): Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 41, 258-290.
- Asch, S. E. & Zukier, H. (1984): Thinking about persons. *Journal of Personality and Social Psychology*, 46:6, 1230-1240.
- Fiske, S. T. & Taylor, S. E. (eds.) (1991): *Social Cognition*. New York: McGraw-Hill.
- Garfinkel, H. (ed.) (1967): *Studies in ethnomethodology*. Englewood Cliffs: Prentice-Hall.
- Goffman, E. (1959): *The presentation of self in everyday life*. New York: Doubleday.
- Goldberg, L. R. (1990): An alternative "Description of personality": The Big-Five factor structure. *Journal of personality and Social Psychology*, 59, 1216-1229.
- Goldberg, L. R. (1992): The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4, 26-42.
- Goldberg, L. R. (1993): The Structure of Phenotypic Personality Traits. *American Psychologist*, 48, 26-34.
- Halasz, F. & Moran, T. P. (1982): Analogy considered harmful. *Human Factor in Computer Systems Conference*, 82, p. 383-6.
- Heider, F. & Simmel, M. (1944): An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243-259.
- Janlert, L.-E. & Stolterman, E. (1997): The character of things. *Design Studies*, 18, 297-314.
- Jones, E. E. (1990): *Interpersonal perception*. New York: Freeman.
- Laurel, B. (1993): *Computers as theatre*. Reading, Mass.: Addison-Wesley.
- McAdams, D. P. (1992): The five-factor model in personality: A critical appraisal. *Journal of Personality*, 60, 329-361.
- Monö, R. (1997): *Design for product understanding: the aesthetics of design from a semiotic approach*. Stockholm: Liber.
- Nardi, B. A. (ed.) (1996): *Context and consciousness – activity theory and human-computer interaction*. Cambridge, Mass: MIT Press.
- Norman, D. A. (1988): *The Design of everyday things*. London: MIT.
- Petti, V. (2000): *Norstedts stora Engelsk-Svenska ordbok : Norstedts comprehensive English-Swedish dictionary*. Stockholm: Norstedts ordbok.
- Reeves, B. & Nass, C. (1996): *The media equation - how people treat computers, television, and new media like real people and places*. Cambridge: Cambridge Univ. Press.
- Saucier, G. (1994): Mini-markers: A brief version of Goldberg's unipolar big-five markers. *Journal of Personality Assessment*, 63, 506-516.
- Saucier, G. (1997): Effects of variable selection on the factor structure of person descriptors. *Journal of Personality and Social Psychology*, 73, 1296-1312.

Schütte, S. (2002): *Designing feelings into products - integrating Kansei Engineering methodology in product development*. Linköping: Linköping studies in science and technology. LiU-Tek-Lic-2002:19.

Sivik, L. & Taft, C. (1992): *Colour combinations and associated meanings: semantic dimensions and colour chords*. Göteborg: Göteborg psychological reports 22(1992):1

Smets, G. J. F. & Overbeeke, C. J. (1995): Expressing tastes in packages. *Design Studies*, 16, 349-365.

Suchman, L. A. (1987): *Plans and situated actions : the problem of human-machine*. Cambridge: Cambridge Univ. Press.

Tabachnik, B. G. & Fidell, L. S. (1996): *Using multivariate statistics*. New York: Harper Collins College Publishers.

Vihma, S. (1995): *Products as representations: a semiotic and aesthetic study of design products*. Helsinki: University of Art and Design Helsinki UIAH.

## APPENDIX A – QUESTIONNAIRE FOR EXPERIMENT 1

Fp \_\_\_\_ Test \_\_\_\_

### Instruktioner

Avsikten med denna studie är att undersöka de uppfattningar som individer bildar sig om olika program. Tänk igenom hur du uppfattade det program du just använde.

Besvara sedan nedanstående enkät som består av 31 motsatta ordpar. Fyll i enkäten genom att med ett kryss markera det beskrivande begrepp på skalan som bäst motsvarar hur du uppfattar att programmet är. De beskrivande begreppen är ”extremt”, ”mycket”, ”något” och ”varken eller”.

### Exempel:

Ordparet är Kraftfull – Kraftlös. Låt säga att du upplevde att programmet var extremt kraftfullt. Då kryssas det beskrivande begreppet ”extremt” på den sida av skalan som ”kraftfull” står.

Extremt   Mycket   Något   Varken  
   eller   Något   Mycket   Extremt

0.            Kraftfull :   X   : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Kraftlös

### Viktigt!

Försök att besvara varje ordpar, om du är osäker på din uppfattning svara med det som instinktivt känns bäst. Du ska bara göra ett kryss per ordpar. Kryssa inom markeringen längs linjen, som i exemplet ovan. Kom ihåg att du är välkommen att fråga om det är något du undrar över.

---

### Enkät

Hur uppfattar du det program som du just tog del av? Beskriv din uppfattning om programmet genom att markera det beskrivande begrepp som bäst motsvarar din uppfattning.

Extremt   Mycket   Något   Varken  
   eller   Något   Mycket   Extremt

1.            Utåtriktad : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Tillbakadragen

2.            Kall : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Varm

3.            Stabil : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Slarvig

4.            Avslappnad : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Nervös

5.            Aningslös : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Klok

6.            Aktiv : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Passiv

Enkäten fortsätter på nästa sida ->

APPENDIX A – QUESTIONNAIRE FOR EXPERIMENT 1 (continued)

	Extremt	Mycket	Något	Varken eller	Något	Mycket	Extremt	
7.								Hämmad : _____ : Vågad
8.								Tillmötesgående : _____ : Krävande
9.								Nonchalant : _____ : Hänsynsfull
10.								Retlig : _____ : Lugn
11.								Förfinad : _____ : Osofistikerad
12.								Overksam : _____ : Verksam
13.								Bestämd : _____ : Osäker
14.								Ovänlig : _____ : Vänlig
15.								Ordentlig : _____ : Oordnad
16.								Fridfull : _____ : Otålig
17.								Ointelligent : _____ : Intelligent
18.								Initiativrik : _____ : Initiativlös
19.								Inåtvänd : _____ : Uppsluppen
20.								Pålitlig : _____ : Opålitlig
21.								Hederlig : _____ : Nyckfull
22.								Besvärlig : _____ : Lättsam
23.								Ytlig : _____ : Djupsinnig
24.								Resignerad : _____ : Vital
25.								Upprymd : _____ : Behärskad
26.								Osympatisk : _____ : Trevlig
27.								Vårdslös : _____ : Aktsam
28.								Avspänd : _____ : Spänd
29.								Okunnig : _____ : Kunnig
30.								Driftig : _____ : Apatisk
31.								Oföretagsam : _____ : Handlande

## APPENDIX B – QUESTIONNAIRE FOR EXPERIMENT 2

Fp \_\_\_\_\_

### Enkät

#### Instruktioner

Avsikten med denna studie är att undersöka de uppfattningar som individer bildar sig om datorprogram. Tänk igenom hur du uppfattade det program du just använde.

Besvara sedan nedanstående enkät som består av 39 motsatta ordpar. Fyll i enkäten genom att med ett kryss markera det beskrivande begrepp på skalan som bäst motsvarar hur du uppfattar att programmet är. De beskrivande begreppen är ”extremt”, ”mycket”, ”något” och ”varken eller”. Alternativet ”varken eller” betyder att programmet av dig uppfattades vara lika mycket av vardera motsatt ord i ordparet.

#### Exempel

Ordparet är **Kraftfull – Kraftlös**. Låt säga att du upplevde att programmet var extremt kraftfullt. Då kryssas det beskrivande begreppet ”extremt” på den sida av skalan som ”kraftfull” står.

Extremt   Mycket   Något   Varken  
   eller   Något   Mycket   Extremt

0.            Kraftfull :   X   : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Kraftlös

#### Viktigt!

Besvara varje ordpar, om du är osäker på din uppfattning svara med det som känns bäst. Gör bara ett kryss per ordpar. Kryssa inom markeringen längs linjen, som i exemplet ovan. Du är välkommen att fråga om det är något du undrar över.

---

Hur uppfattar du det program som du just använde? Beskriv din uppfattning om programmet genom att markerade det beskrivande begrepp som bäst motsvarar din uppfattning.

Extremt   Mycket   Något   Varken  
   eller   Något   Mycket   Extremt

1.            Utåtriktad : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Hämmad
2.            Kall : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Varm
3.            Stabil : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Slarvig
4.            Avslappnad : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Nervös
5.            Aningslös : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Klok
6.            Aktiv : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Passiv
7.            Tillbakadragen : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : \_\_\_\_\_ : Vågad

**APPENDIX B – QUESTIONNAIRE FOR EXPERIMENT 2 (continued)**

		Extremt	Mycket	Något	Varken eller	Något	Mycket	Extremt	
8.	Tillmötesgående	:	:	_____	:	:	:	:	Krävande
9.	Nonchalant	:	:	_____	:	:	:	:	Hänsynsfull
10.	Retlig	:	:	_____	:	:	:	:	Lugn
11.	Förfinad	:	:	_____	:	:	:	:	Dum
12.	Overksam	:	:	_____	:	:	:	:	Verksam
13.	Bestämd	:	:	_____	:	:	:	:	Osäker
14.	Ovänlig	:	:	_____	:	:	:	:	Vänlig
15.	Ordentlig	:	:	_____	:	:	:	:	Oordnad
16.	Fridfull	:	:	_____	:	:	:	:	Otålig
17.	Ointelligent	:	:	_____	:	:	:	:	Intelligent
18.	Initiativrik	:	:	_____	:	:	:	:	Initiativlös
19.	Inåtvänd	:	:	_____	:	:	:	:	Uppsluppen
20.	Pålitlig	:	:	_____	:	:	:	:	Opålitlig
21.	Hederlig	:	:	_____	:	:	:	:	Nyckfull
22.	Besvärlig	:	:	_____	:	:	:	:	Lättsam
23.	Ytlig	:	:	_____	:	:	:	:	Insiktsfull
24.	Resignerad	:	:	_____	:	:	:	:	Vital
25.	Upprymd	:	:	_____	:	:	:	:	Behärskad
26.	Osympatisk	:	:	_____	:	:	:	:	Trevlig
27.	Vårdslös	:	:	_____	:	:	:	:	Aktsam
28.	Avspänd	:	:	_____	:	:	:	:	Spänd
29.	Okunnig	:	:	_____	:	:	:	:	Kunnig
30.	Driftig	:	:	_____	:	:	:	:	Apatisk
31.	Blyg	:	:	_____	:	:	:	:	Entusiastisk
32.	Överseende	:	:	_____	:	:	:	:	Dominant
33.	Harmonisk	:	:	_____	:	:	:	:	Kaotisk
34.	Trygg	:	:	_____	:	:	:	:	Otrygg
35.	Reserverad	:	:	_____	:	:	:	:	Lekfull
36.	Energisk	:	:	_____	:	:	:	:	Foglig
37.	Återhållsam	:	:	_____	:	:	:	:	Livlig
38.	Oföretagsam	:	:	_____	:	:	:	:	Handlande
39.	Osäker	:	:	_____	:	:	:	:	Säker