What is beautiful is usable

Article *in* Interacting with Computers · December 2000

DOI: 10.1016/S0953-5438(00)00031-X · Source: DBLP

3 authors, including:

Adi Katz
Shamoon College of Engineering
17 PUBLICATIONS **756** CITATIONS

ELSEVIER

Interacting with Computers 13 (2000) 127–145

Interacting with Computers

www.elsevier.com/locate/intcom

What is beautiful is usable

N. Tractinsky*, A.S. Katz, D. Ikar

Industrial Engineering and Management, Ben Gurion University of the Negev, Beer Sheva 84105, Israel

Received 10 January 1999; revised 2 March 2000; accepted 24 March 2000

Abstract

An experiment was conducted to test the relationships between users' perceptions of a computerized system's beauty and usability. The experiment used a computerized application as a surrogate for an Automated Teller Machine (ATM). Perceptions were elicited before and after the participants used the system. Pre-experimental measures indicate strong correlations between system's perceived aesthetics and perceived usability. Post-experimental measures indicated that the strong correlation remained intact. A multivariate analysis of covariance revealed that the degree of system's aesthetics affected the post-use perceptions of both aesthetics and usability, whereas the degree of actual usability had no such effect. The results resemble those found by social psychologists regarding the effect of physical attractiveness on the valuation of other personality attributes. The findings stress the importance of studying the aesthetic aspect of human–computer interaction (HCI) design and its relationships to other design dimensions.

Keywords: User interface; Aesthetics; Usability; Apparent usability; HCI design considerations; HCI perceptions

1. Introduction

The tension between form and function has long been at the crossroad of artifact design [1–3]. Whereas emphasis on function stresses the importance of the artifact's usability and usefulness, accentuating the artifact's form serves more the aesthetic, and perhaps social, needs of designers and customers. Until the first quarter of this century, the design of commodities and mass production artifacts were quite devoid of aesthetic considerations. Petroski [3] credits two industrial design pioneers, Loewy and Dreyfuss, with the introduction of aesthetic considerations to mass production and with the development of industrial design as an explicit marketing instrument. Evidently, aesthetics considerations gained importance quickly. About half a century later, Norman [4,5] laments the appropriation of modern design by designers who place aesthetics ahead of usability. Similar sentiments concerning designers' priorities can be found in various areas of artifact design.

* Corresponding author.

0953-5438/00/$ - see front matter
PII: S0953-5438(00)00031-X

For example, Tufte [6] denounces the increasing use of "chartjunk"—the application of unnecessary decoration and ink elements that hinder the efficient reading of visual displays.

Perhaps in a backlash to recent tendencies by the computer industry to oversell glitz and fashion in its products or because of its origins in disciplines that emphasize efficiency, the field of human-computer interaction (HCI) appears to stress the prominence of usability over aesthetics. There is little doubt that, in general, the criterion of aesthetic design is an integral part of effective interaction design [7]. Yet, readers of HCI textbooks can hardly find any reference to aesthetic considerations in design. Contrast this with the plethora of publications that deal with the concept of usability and its application and evaluation. The claim for the prominence of usability in the field of HCI goes sometimes as far as stating that "measurement of usability defines the success or lack thereof in a GUI design" [8, p. 60].

In a sense the concepts of aesthetics and usability represent two orthogonal dimensions of HCI. Whereas aesthetics usually refers often to non-quantifiable, subjective, and affect-based experience of system use, usability is commonly measured by relatively objective means and sets efficiency as its foremost criterion [8].

The near neglect of the aesthetic aspect of HCI is unfortunate for several reasons. First, it reveals a gap between the practice of much of the computer industry and the research foci of HCI in terms of the attention given to HCI design criteria. Second, it seems to ignore important needs of computer users, who, like consumers of other commodities [9,10] are likely to value aesthetics and fashionable designs in addition to usability. Third, previous research suggests that aesthetic perceptions of an interface are highly correlated with perceptions of the interface's ease of use [11,12]. Therefore, it appears that users do not perceive these two design dimensions as independent. Thus, our purpose in this study is to experimentally explore users' perceptions of usability and aesthetics before and after using an application, in order to tease out how they effect users' evaluation of the quality of their interaction with the application.

2. Background

Much effort has been put in recent decades into studying, and then advocating, the effective design of HCI. One of the central sub-disciplines that emerged in the field during the early 1980s was that of usability engineering [13,14]. With a few exceptions (e.g. [15], who emphasized the user's experience), usability engineering has traditionally emphasized objective performance criteria, such as time to learn, error rate and time to complete a task [8]. Recently, as usability became one of HCI's most popular concepts [16,17], less stringent and more subjective criteria such as user satisfaction have also been widely recognized as appropriate measures of good HCI design [18,19].

There is a pervasive claim for the prominence of usability criterion over aesthetics in HCI. Even when classical elements of aesthetics—such as screen design and graphics—are dealt with [20–22], they are mostly analyzed in terms of their effects on human information processing rather than on human affect and experience. Whenever considerations of aesthetics and usability contradict, the unequivocal recommendation is to give

priority to the latter [20,21]. This sentiment is expressed today with regard to the design of Web sites as well. For example, in a guide to Web site usability, Spool et al. [23] argue that "… no one surfs the online employee policy manual just for kicks" (p. 4). They conclude "we found no evidence that graphic design helps users *retrieve information* on a web site" (p. 83, italics added). Unfortunately, this approach ignores the affective aspect of users' behavior [24] and its relevance to HCI. It also ignores the potential relationships that exist between usability and aesthetics and their combined effect on users' evaluation of computer systems. For example, Jordan [10] found that usability and aesthetics (as well as several other factors) are both instrumental in creating pleasurable electronic products.

Because the aesthetics aspect was by large overlooked in HCI research, it is not surprising that only a handful of studies to date have investigated the relations between users' perceptions of aesthetics and usability. Kurosu and Kashimura [11] explored the relationships between *a priori* perceptions of the ease of use of an automatic teller machine (ATM)—which they termed "apparent usability"—and other variables. These variables included factors believed by HCI professionals to enhance usability (termed "inherent usability" by Kurosu and Kashimura). Another factor included in Kurosu and Kashimura's study was the perceived beauty of the interface. The study, conducted in Japan, found surprisingly high relationships between users' judgements of the interface's aesthetics and its apparent usability $r = 0.59$). (In fact, this correlation was higher than all but one of the inherent usability factors.) In a recent study, Tractinsky [12] corroborated Kurosu and Kashimura's findings in a different culture (Israel), and removed doubts about potential method bias as an alternative explanation to the high correlation between apparent usability and perceptions of interface aesthetics.

The mechanism that links affective and cognitive evaluations of user interfaces is not clear. We speculate that the strong correlations found between perceived usability and aesthetics resemble findings in the social psychology literature about the relationships between physical attractiveness and socially desirable characteristics. This conjecture is in line with the recent evidence suggesting that individuals experience media similar to their experience of social environments and interactions. Reeves and Nass [25] presented the results of an extensive research project, which found that "individuals' interactions with computers, television, and new media are *fundamentally social and natural*, just like in real life" (p. 5, original emphasis). The social phenomenon of inferring personality attributes from physical attractiveness was demonstrated by Dion et al. [26]. In a paper titled "What is beautiful is good," Dion et al. found that people who are physically attractive are assumed (by other people) to possess more socially desirable personality traits than persons who are unattractive. Researchers suggested two mechanisms by which the "beautiful is good" phenomenon can be explained. The first mechanism suggests that the carry over from the physical appearance to other traits may reflect a stereotyping approach, which associates beauty with other personal attributes [26,27]. The other explanation involves the halo effect. That is, because physical beauty is the most obvious and accessible personal characteristic accessible to others [26], it is perceived early in the interaction and then tends to color later perceptions and inferences about other personal characteristics.

Recently, issues of aesthetics, and more generally, affective properties of the shopping environment became the foci of research in the fields of marketing and consumer behavior.

The concept of "retail personality" was introduced by Martineau [28], who illustrated how a store's functional and psychological characteristics are capable of defining its image in a consumer's mind. Russell and Pratt [29] suggested that a store's affective quality (which is based to a large extent on its physical characteristics, including aesthetics) must be considered in any complete description of the shopping environment. Donovan and Rossiter [30] found that feelings generated by a retail store environment relate to shopping behavior while Bloch [31, p. 16] concluded that the "physical form or design of a product is an unquestioned determinant of its marketplace success". Edell and Burke [32] demonstrated that feelings generated by advertisements (especially by the nonverbal elements of the ad) may subsequently influence cognitive processing of the ad's content. A study by Darden and Babin [9] found that not only are store characteristics capable of enticing affective response in consumers, but that a store's functional (i.e. more objective) and affective (i.e. subjective) qualities are interrelated in the consumers' eyes. Similar to the "beauty is good" phenomenon in social psychology, consumer researchers suspect that a halo effect is responsible for a carry over of first impressions of products or shopping environments to consumers' evaluations of other attributes of these products or environments [33]. In addition, this literature suggests that aesthetics may affect perceptions of products by inducing affective response which, in turn, influence evaluations of other product attributes and of the product in general. In Fazio et al., [34, p. 212] terms, "affect is preattentively 'extracted' and influences subsequent perception.".

To conclude, three different processes may induce positive relationships between interface aesthetics and perceived usability: (a) A popular stereotyping which might associate successful design on one (noticeable) design dimension with successful design of other, less implicit design dimensions. (b) A halo effect may cause carry over of an aesthetic (or not aesthetic) design to perceptions of other design features. (c) An affective response to the design's aesthetics may improve users' mood and their overall evaluations of the system.

3. Research questions

Our first goal in this study was to test whether the initial correlation of perceived aesthetics and usability reflect more general tendencies to associate aesthetics with other system attributes. The question is important, because a positive answer would portray a picture of users whose judgment is completely masked by the interface's aesthetics. If we find that users are able to discriminate between various system attributes and their relations to the interface's aesthetics, we would have to conclude that perceptions of aesthetics and usability represent a special type of relationships that merit further study. Again, one can find similarity here with the history of research on the "beautiful is good" phenomenon. As the study of physical attractiveness' effects on personality judgments evolved, it was shown that these effects do not apply universally to all personality attributes. Individuals were found to be selective in their association of physical attractiveness and personal qualities. In a comprehensive analysis of research in the field, Eagly et al. [27] found that the "beautiful is good" holds strongly especially for social competence attributes. These relationships were less strong for other personality traits (e.g. intellectual

competence), and were demonstrated to disappear for yet a different type of attributes, (e.g. integrity).

Our second goal is to explore what happens to users' perceptions of aesthetics and usability *after* they use the system. While Kurosu and Kashimura [11] and Tractinsky [12] established the existence of close relationships between perceptions of aesthetics and apparent usability *before* users actually use the system, it still remains to be seen whether users continue to perceive these concepts as highly correlated after they interact with the system. Dearth of previous research limits our ability to hypothesize about the perseverance of these relationships. On the one hand, intuition suggests that post-use perceptions of usability are highly dependent on the interface's actual usability rather than on other features. On the other hand, social psychology research found that initial social perceptions persevere even after strong evidence is presented to the contrary [35,36]. In a study of information systems use, Hiltz and Johnson [37] found that initial perceptions of a system's ease of use were correlated with users' satisfaction with the system's interface after four months of use. Similarly, Szajna and Scamell [38] and Gaeth et al. [39] found that users' expectations of a system affects their perceptions of the system's actual performance. Consequently, if initial perceptions of usability are highly related to perceptions of interface aesthetics, these relations may hold even after experiencing the interface's actual usability. Thus, our study is designed to test whether the initial perceptions of aesthetics-usability relationships hold after a period of system use, and whether these perceptions are affected by the degree of the interface's perceived aesthetics and/or by the actual usability of the system.

4. Method

4.1. Participants

Participants were 132 third year Industrial Engineering (IE) students who participated in the experiment for class credit. The IE curriculum includes primarily technically oriented courses, with no product design or aesthetic components. The participants were not exposed during their academic studies prior to the experiment to aesthetic considerations of artifact design. The average age of the participants was 25. About 67% of the participants were males.

4.2. Experimental design and manipulations

The experiment used a 2×2 between groups factorial design. One factor, the aesthetic level of the interface, had three levels: low, medium and high. The other factor, usability, included two levels: low and high.

The Aesthetics Factor. In the first stage of the experimental session, a computer program presented the participants with nine ATM layouts. The layouts were selected from the 26 ATM layouts that were introduced by Kurosu and Kashimura [11] and later adapted by Tractinsky [12]. The nine layouts were chosen based on the ratings of the screens by participants in Tractinsky's study [12]. Three of the nine layouts were rated as highly aesthetic in that study, three layouts were rated low in terms of aesthetics and the other

three were rated in between. All of the layouts contained the same objects, and differed only in terms of the way those objects were arranged on the screen. Participants were asked to rate each layout on a 1–10 scale with regard to three attributes. One of the attributes was the layout's aesthetics. Participants who were assigned to a certain aesthetics condition (high, medium, or low) performed the experimental tasks using a layout that matched that condition based on their ratings. For example, participants who were assigned to the high aesthetics condition worked with the layout that they rated as the most aesthetic. To ensure successful manipulation of the aesthetics factor, eight participants whose ratings of the most- and least-aesthetic layouts differed by 3 points or less (on a 1–10 scale) were excluded from the analysis.

The Usability Factor. The program presented the participants with 11 tasks to be performed on the ATM. The tasks (detailed under the Tasks subsection) were presented in a constant order to all participants. The usability factor was manipulated by introducing difficulties to the interaction between the ATM and the participants in the low usability condition. These difficulties included longer system delays (of 9 s on average per task), buttons that did not operate the first time they were pressed, and one task that prevented the users from taking a short cut for its completion. To sensitize participants to the ATM's usability, they were told that they should complete each task as quickly as possible.

4.3. Procedure

The experimental session included three stages. In the first stage, the experimental program displayed the nine ATM layouts. Each layout was presented three times. Every time a layout was presented, participants were asked to rate it on one of three dimensions: (i) aesthetics; (ii) ease of use; and (iii) amount of information on the screen. Thus 27 combinations of layout (9) on rating (3) were presented in a completely randomized order. An example of an ATM layout with the rating question and response buttons is presented in Appendix A.

Before the second stage, participants were assigned to an aesthetics condition as described above. They then practiced the use of the ATM by performing the four types of tasks that they later had to perform in the experiment. After practicing the ATM, participants in each aesthetic level were randomly assigned to one of the two usability levels, and performed the experimental tasks using the same layout that they used for the practice tasks. Finally, participants were asked to rate the system with which they worked on several dimensions.

4.4. Tasks

Participants received a 4-digit personal code that allowed them to operate the ATM. The code was, in fact, identical for all participants. Participants had to perform 11 tasks which were comprised of the following four types: inquiring about their account balance (this task type was repeated three times during the experiment); withdrawing cash (four times using different amounts of cash); checking out the account balance and withdrawing cash simultaneously (twice); and depositing money (twice). Obviously, all tasks were performed logically without the exchange of physical materials. The tasks were presented at the bottom of the screen, in an area separated from the ATM display, one task at a time.

Table 1
Pre-experiment mean ratings (standard deviations in parentheses) of aesthetics, usability and amount of information for the ATM interface that was later used in the experiment in each experimental group. Rating scales ranged from 1 (low) to 10 (high)

Aesthetic level	Pre-experimental perceived measure	Usability	
		High	Low
High	Aesthetics	8.48 (1.25)	8.05 (.94)
	Usability	7.62 (1.53)	6.90 (1.55)
	Information	4.91 (1.48)	4.80 (1.99)
	N	21	20
Medium	Aesthetics	5.05 (1.05)	4.90 (.99)
	Usability	5.20 (2.17)	3.84 (2.43)
	Information	5.45 (1.95)	5.63 (1.54)
	N	20	19
Low	Aesthetics	2.13 (1.1)	2.0 (1.22)
	Usability	4.04 (2.23)	3.19 (2.23)
	Information	5.61 (1.53)	6.57 (1.43)
	N	23	21

The main panel of the ATM display presented system messages that guided the users in performing their tasks (e.g. "Enter your personal code"), or feedback about the task progress (e.g. "the system is handling your request, please wait"). After successfully completing a task, participants could move to the next task at their own discretion. An example of the screen in one task, querying the account balance, is presented in Appendix B.

4.5. Variables

The variables of interest in this study are subjective valuations of interface properties. Three variables measured pre-experimental perceptions of the interface: aesthetics, usability, and the amount of information it contains. Four variables measured post-experimental perceptions of the interface. Three of these measures correspond to the three pre-experimental variables (aesthetics, usability and amount of information). The fourth variable measures user's satisfaction. Because of the nature of our study (in which participants had to answer the same items for each of the nine displays before the main experimental procedure) we decided to settle for single-item measures.[1] In addition to subjective measures, number of errors and task completion times were recorded for each task.

[1] We are well aware that measurement theory advocates the use of multiple item measures. However, as Wanous et al. [50] suggest, single-item measures can be accepted under circumstances similar to those of our study (e.g. when there is danger that too many items might overwhelm the respondents), provided that the items are focused and unambiguous.

Table 2
A correlation matrix of pre-, and post-experimental measures ($n = 124$). The dotted-lines separate pre-experimental correlations between three measures (top-left), post-experimental correlations (bottom-right), and correlations between pre-, and post-experimental measures (top-right) (*$p < 0.01$ level)

	Pre-usability	Pre-information	Post-aesthetics	Post-usability	Post-information	Post-satisfaction
Pre-aesthetics	0.66*	−0.26*	0.62*	0.50*	−0.14	0.48*
Pre-usability	–	−0.18	0.52*	0.48*	−0.11	0.48*
Pre-information		–	−0.03	0.11	0.63*	0.00
Post-aesthetics			–	0.71*	−0.02	0.71*
Post-usability				–	−0.01	0.87*
Post-information					–	−0.10

5. Results

5.1. Manipulation check

Table 1 displays the *pre-experimental* mean ratings of the three interface dimensions (aesthetics, usability and amount of information). These ratings refer to the one ATM interface that was eventually used in the experiment by the participants in each experimental condition. A one-way analysis of variance (ANOVA) revealed a strong effect of the aesthetics factor ($F(2, 121) = 338.58;\ p < 0.001$). Mean ratings of the ATM's aesthetics were 8.26, 4.97 and 2.07 for the high, medium, and low aesthetics conditions, respectively. To test whether the differences between the three conditions were statistically significant, we used the standard method of Scheffe post hoc contrasts [40]. The differences between any pair of three groups were significant at the 0.001 level, indicating that the aesthetics manipulation was successful.

The success of the usability manipulation was evaluated by comparing the completion times of the 11 tasks for both usability conditions. We performed a 2-way ANOVA to find usability and aesthetic effects on completion times. The usability factor had significant effect on completion times ($F(1,\ 118) = 414.05;\ p < 0.001$), indicating that the usability manipulation succeeded. In the high-usability condition the average completion time was 23.5 s per task, whereas in the low-usability condition the average was 37.5 s. Of this 14 s average difference, 9 s were caused by intentional system delays of the experimental program, while the other 5 s can be largely attributed to other usability differences (as noted in Section 4). At the same time, there was no aesthetics effect on completion times ($F(2,\ 118) = 0.01;\ p = 0.987$), nor was there an aesthetics X usability interaction effect ($F(2, 118\ ) = 0.96;\ p = 0.387$). Thus, we conclude that the usability manipulation was successful and clear of any aesthetics side effects.

On average, participants made 1.26 mistakes during the experimental sessions. There were neither main, nor interaction effects on the numbers of errors.

5.2. Correlation analysis

Intercorrelations among the perceived measures in this study (both before and after the experiment) are presented in Table 2. Pre experimental perceptions of ATM aesthetics and their perceived usability were highly correlated $r = 0.66$). The high correlation resembles those obtained by Kurosu and Kashimura [11] and Tractinsky [12]. Perceived amount of information, on the other hand was only weakly, and negatively, correlated with the perceived aesthetics $r = -0.26$) and perceived usability $r = -0.18$). This indicates that participants were able to distinguish between these concepts. It further corroborates Tractinsky's [12] conclusion that it is unlikely that the correlation between aesthetics and apparent usability stems from a method effect that creates artificial correlations between the interface's attributes. The correlations between perceived aesthetics and usability remained high $r = 0.71$) even after the experiment. In addition, post-experimental satisfaction with the ATM was highly correlated not only with its perceived usability (as predicted by the usability literature), but also with post-experimental perceptions of aesthetics ($r = 0.87$ and.71, respectively). Pre-, and post-experimental correlations of

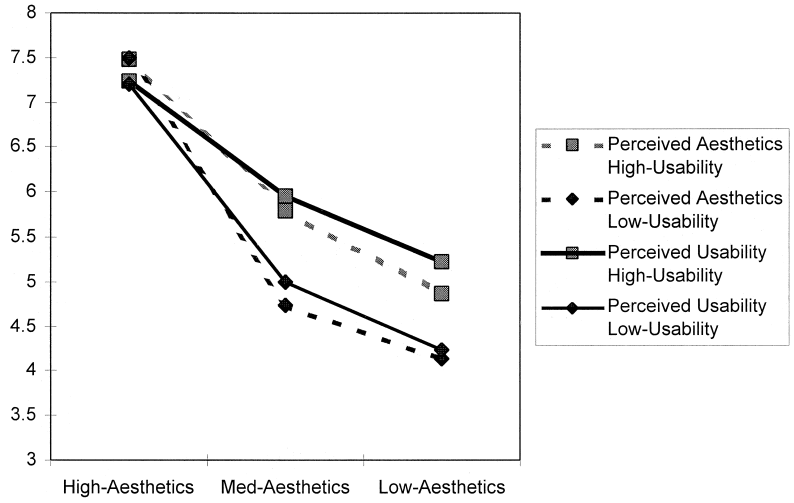


Fig. 1. Post-experimental perceptions of usability and aesthetics (on a 1–10 scale) under three levels of ATM aesthetics and two levels of ATM usability.

aesthetics and amount of information were relatively high ($r = 0.62$ and 0.63, respectively), while pre-, and post-experimental correlation of perceived usability was lower $r = 0.48$). The relatively low correlation between the two usability measures can be explained in part by the different experience that the two usability groups had using the ATM in this experiment. (The correlation was 0.53 for the high-usability group, and 0.40 for the low-usability group.) Another interesting finding is that post-use satisfaction could be predicted quite well from pre-experimental perceptions of usability and of aesthetics $r = 0.48$).

5.3. Analysis of variance

The post-experimental measures of perceived aesthetics and usability are presented in Fig. 1 for the six experimental conditions. The results indicate a very similar pattern of both measures, which decline considerably as the level of aesthetics decrease. There is also a slight tendency for those measures to decrease under low usability, but only under moderate and low levels of aesthetics. Under the high-aesthetics condition, the perceptions of both usability and aesthetics measures were insensitive to changes in the ATM's usability.

The main analysis of the experimental results was done using multivariate analysis of covariance (MANCOVA). The use of this technique is recommended when a variable (i.e. the covariate) which is likely to affect the dependent variables is also assumed to be correlated with one (or more) of the experimental factors [41]. By applying this technique, the researcher can test the experimental effects as if all participants scored the same on the covariate. Otherwise, the researcher may not be able to tell whether observed differences in the dependent variable stem from differences in the experimental factor or from differences in the covariate. The MANCOVA's two factors were level of Aesthetics (3 levels) and Usability (2 levels). Post-experiment ratings of aesthetics, usability, amount of

Table 3
Results and significance levels of univariate and stepdown F-tests of the effects of the Aesthetics and the Usability factors on post-experimental measures, with pre-experimental perceptions of usability as a covariate (*$p <$ 0.01; **$p < 0.001$)

Factor	Dependent variable (post-experimental perceived measures)	Univariate F (df)	Stepdown F (df)
1. Covariate (pre-exp. perceived usability)	Usability	7.64* (1, 117)	7.64* (1, 117)
	Aesthetics	7.76* (1, 117)	2.04 (1, 116)
	Satisfaction	7.02* (1, 117)	0.10 (1, 115)
	A. of information	0.16 (1, 117)	0.37 (1, 114)
2. Aesthetics	Usability	4.75* (2, 117)	4.75* (2, 117)
	Aesthetics	9.73** (2, 117)	4.49* (2, 116)
	Satisfaction	4.88* (2, 117)	0.06 (2, 115)
	A. of information	0.92 (2, 117)	1.22 (2, 114)
3. Usability	Usability	1.38 (1, 117)	1.38 (1, 117)
	Aesthetics	1.17 (1, 117)	0.25 (1, 116)
	Satisfaction	2.92 (1, 117)	1.47 (1, 115)
	A. of information	0.04 (1, 117)	0.01 (1, 114)
4. Interaction (aesthetics by usability)	Usability	0.66 (2, 117)	0.66 (2, 117)
	Aesthetics	0.73 (2, 117)	0.31 (2, 116)
	Satisfaction	0.94 (2, 117)	1.58 (2, 115)
	A. of information	0.01 (2, 117)	0.06 (2, 114)

information and satisfaction were the dependent variables. Pre-experiment ratings of apparent usability served as a covariate in the analysis because they were correlated with the aesthetics factor and with the dependent variables.[2] The statistical software used was SPSS 6.1.3.

Before proceeding with the analysis, we certified that the data fulfilled the major assumptions of MANCOVA. Normality of the dependent variables was tested for each experimental group. None of the distributions' skewness and kurtosis was statistically significant below the recommended threshold of 0.01 [42, p. 214]. Box's multivariate test for homogeneity of the covariance matrices was insignificant at the 0.05 level ($M = 70.44$; $F(50, 24943) = 1.28$; $p = 0.09$). Similarly, an overall test of regression homogeneity (parallel slopes) was insignificant ($F(20, 362.46) = 1.32$; $p = 0.16$) as well as for each of the stepdown tests [41].

Because of the relatively high correlations between some of the dependent variables, we used the Roy–Bargmann stepdown analysis (p. 403 and 328 in Refs. [41,42], respectively). Since there was no theoretical basis for the order in which the dependent variables

[2] Note that there was no need to use pre-experimental perceptions of aesthetics as a covariate for the usability factor. Since participants in one usability condition are actually matched against participants in the other usability condition in terms of their pre-experimental ratings of aesthetics, these two variables are not correlated.

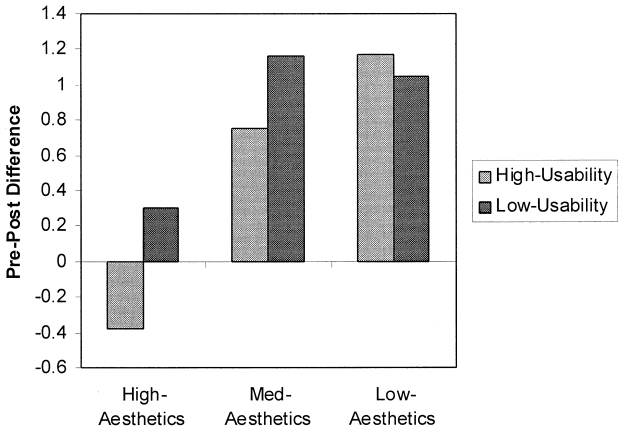


Fig. 2. Differences between post-experimental and pre-experimental perceptions of the ATM's aesthetics.

should enter the analysis, post-experimental usability was entered first, being the main variable of interest in this study. Table 3 shows the results of univariate F-tests and Roy–Bargmann Stepdown F-tests regarding the effects of the covariate and the experimental factors on the four dependent variables.

After adjusting the pre-experimental level of perceived usability (Row 1 in Table 3) the stepdown MANCOVA results clearly indicate that the only factor affecting post-experimental measures was Aesthetics (see Row 2 in Table 3). The level of aesthetics affected not only the post-experimental perceptions of the ATM's aesthetics (stepwise $F(2, 117) = 4.75$; $p < 0.01$), but the post-experimental perceptions of its usability as well (stepwise $F(2, 116) = 4.49$; $p < 0.01$). Post-experimental satisfaction, which was associated with level of aesthetics in the univariate analysis (univariate $F(2, 117) = 4.88$; $p < 0.01$), was not significant in the stepdown analysis (stepwise $F(2, 115) = 0.06$). That is, the apparent effect of the Aesthetics factor on satisfaction is attributed to the fact that perceived satisfaction is correlated with perceived aesthetics and usability rather than to a unique effect that the Aesthetics factor had on satisfaction. Table 3 also shows (Row 3 in Table 3) that there was no effect of the usability factor on any of the post-experimental perceptions. Similarly, the Aesthetics and the Usability factor did not interact to affect post-experimental perceptions (Row 4).

Next, we performed contrasts of perceived aesthetics and usability between the three levels of aesthetics, using pre-experimental perceptions of usability as a covariate. Perceived aesthetics was significantly different between the high-aesthetics condition and the other two conditions ($t = 4.33$; $p < 0.001$ between the high-aesthetics and the low-aesthetics conditions, and $t = 3.42$; $p < 0.001$ between high-aesthetics and medium aesthetics). The difference between the low-aesthetics and the medium aesthetics groups was not statistically significant. Contrasts of perceived usability among the aesthetics conditions found similar results ($t = 3.06$; $p < 0.01$ between the high-aesthetics and the low-aesthetics conditions, and $t = 2.20$; $p = 0.03$ between high-aesthetics and medium aesthetics).

The results thus far relate to post-experimental perceptions. Next, we tested whether the

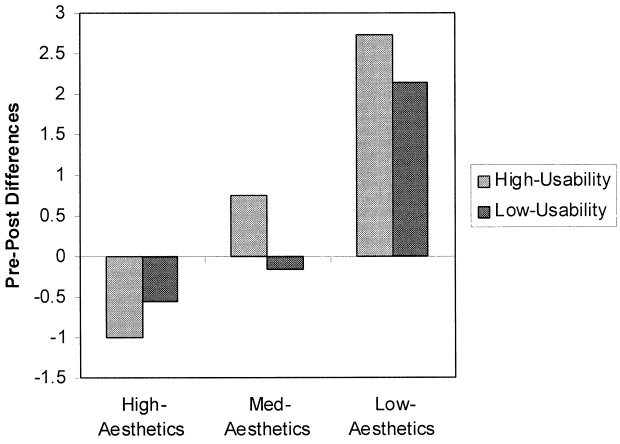


Fig. 3. Differences between post-experimental and pre-experimental perceptions of the ATM's usability.

experimental manipulations affected how participants *changed* their perceptions of the ATM following the experiment. The change in perceptions for each participant was defined as the difference between the post-experimental and the pre-experimental measures of perceived aesthetics, usability, and amount of information. We ran a 3 (aesthetics) $\times$ 2 (usability) ANOVA with the three change measures as dependent variables. Fig. 2 present the data for the differences in perceptions of aesthetics. For the change measure in aesthetics perceptions there was a significant effect of the aesthetics factor ($F(2, 118) = 36.04$; $p < 0.001$). Surprisingly, this effect is associated with a decrease in perceptions of the interface's aesthetics in the high-aesthetics condition, and an increase in aesthetic perceptions in the medium and the low aesthetics group. Scheffe's post hoc tests were significant at the 0.05 level between all three aesthetics levels. There were no usability main effects or a usability $\times$ aesthetics interaction with regard to the change score in aesthetics perceptions.

The same analysis was conducted for differences in perceptions of usability before and after the experiment. The results indicate weak aesthetics main effect ($F(2, 118) = 2.63$; $p = 0.07$), but no usability or interaction effects. Apparently, the effect of level of aesthetics on the changes in perceptions of usability was similar to the changes in perceptions of aesthetics. There was a positive change in perceptions of usability as the level of aesthetics decreased (see Fig. 3).

6. Discussion

This study corroborates the results of earlier studies [11,12] that found strong correlations between users' perception of an interface aesthetics and their perception of the usability of the entire system. We also demonstrated that users are capable of distinguishing between various properties of the system. For example, in our study users did not

associate the interface's attribute of amount of information with the interface's beauty or ease of use. This finding reinforces the claim that the aesthetics-usability association is a genuine phenomenon and not the result of an evaluation method bias [12]. Our study was not designed to explore the process that leads to this association, thus further studies are needed to shed more light on the cognitive and/or affective processes that lead users to associate interface aesthetics with other system attributes.

Perhaps the most interesting finding of this study, though, relates to the question of what most strongly determine post-use perceptions of usability, aesthetics and overall satisfaction with the system. An intuitive answer to this question would probably be that aesthetics affects perceptions of aesthetics, actual usability affects perceptions of usability, and that some combination of the two affects satisfaction with the system. The finding that the participants who interacted with an aesthetic ATM perceived the system as more aesthetic than participants who interacted with less aesthetic ATMs is therefore expected. However, most surprising is the fact that post-experimental perceptions of system usability were affected by the interface's aesthetics and not by the actual usability of the system.

While overall evaluations of aesthetics and usability were higher when using the more aesthetic interface, we also found an interesting finding regarding changes between the pre- and the post-experimental evaluations. Users in the medium- and low-aesthetics groups tended to evaluate the ATM's usability and especially its aesthetics more favorably after they used it relative to their initial evaluations. These results might be explained in terms of a "mere exposure" effect [43], which suggests that our evaluations of an object improve after we are repeatedly exposed to that object. However, as Judd and Brauer [44] indicate, there seem to be several processes that operate during repeated exposures to an object, some of which lead to even lower evaluations of object that were initially evaluated negatively. Perhaps, these results reflect the process of users' adapting to a system with which they had to interact. In a sense it is the HCI version of "love the one you're with". Clearly, more research is needed on this issue.

This research was motivated primarily by the tension between function and form in HCI. The results suggesting that interface aesthetics has a major effect on a priori perceptions of ease of use, and perhaps more importantly on post facto evaluations of usability may come as an uneasy surprise to those versed in the field of HCI. The view of usability and aesthetics as interrelated attributes questions the unequivocal message expressed by the HCI literature in favor of the former over the latter [20,21].

The design of information systems in general and of the user interface specifically is often compared to architectural design. In noting the potential similarity between these two design disciplines, Hooper [45, p. 13] wrote that architectural analyses emphasize the facades of buildings for several reasons. Among those reasons are that the facade is the introduction to the building: "This is what most people experience directly". In addition, the facade can serve as a "membrane between the inside and the outside, and… its purpose is to articulate the relationship between the two." However, Hooper later dismisses the viability of the "interface as facade" approach. Our study demonstrates that regardless of its desirability to interface designers, users might relate to this analogy. The facade of an information system is what users experience first and it is what cues users about the inside. Moreover, the facade taints how the user perceives further interactions with the system.

The implications for interface design are clear. The fact that users perceive aesthetically

appealing interfaces as indicative of usable systems calls for an integrative approach to interface design which will take simultaneous account of the two seemingly unrelated properties. Intuitively, it would seem that the importance of aesthetics design is relevant mainly to systems that are used voluntarily. There is a higher likelihood of purchasing or using a system that is perceived to be more usable (as well as more aesthetic). However, the results of this study might be relevant even in situations where users are compelled to use a system. For one, as shown here, there are strong correlations between users' satisfaction from using the system and their perceptions of its aesthetics and usability (which by itself is affected by perceived aesthetics). User satisfaction is important even for involuntary system use. Second, the advantage of aesthetic interfaces might transcend the mere (yet important) *perceptions* of the system's quality to the domain of actual performance. As demonstrated by Isen [46], positive affect is likely to improve decision making and creativity. Thus, the positive affect created by aesthetically appealing interfaces may be instrumental in improving users' performance as well.

We should note that this study has certain limitations, as is the case with any exploration of new research venues. The post-use evaluations of the ATM's attributes in our experiment were obtained after a relatively short (though quite intense) time period. The results thus, are only indicative of a short-term influence of aesthetic perceptions. Clearly, more research is needed to establish its long-term influence. In addition, this study presents only bottom-line data (that is, only the end results of the human–computer interaction). Thus, the *process* underlying the relationships between aesthetics, perceived usability and other system attributes can only be speculated. Better understanding of these relationships requires the use of more process-oriented data. It should also be pointed out that the relative homogeneous nature of the participants in this study (all were engineering students) restricts the generalizability of the experimental results. Obviously, generalization could have been broader if participants had more heterogeneous backgrounds. However, given that engineering students with broad computer experience are likely to represent a relatively rational approach towards information technology, the current results may look even more striking. In a sense, one would be hard pressed to find a different group of students who could better distinguish between the appearance and the behavior of an artifact. On the other hand, it might be that more mature participants (e.g. professionals) would have perceived the aesthetics-usability relations differently. Thus, it is recommended that future studies indeed sample from such populations.

In addition to further exploration of the nature of relationships between interface aesthetics and users' perceptions of other system attributes, one of the most intriguing questions is: What makes an aesthetic interface? A few general guidelines exist already in the field of HCI [21] and in influential works on esthetical preferences in general [47]. Yet, as Martindale et al. [48] demonstrated, these guidelines and other commonly held beliefs do not always withstand empirical scrutiny. Again, work in the area of marketing and consumer behavior [49] may provide initial basis for such research.

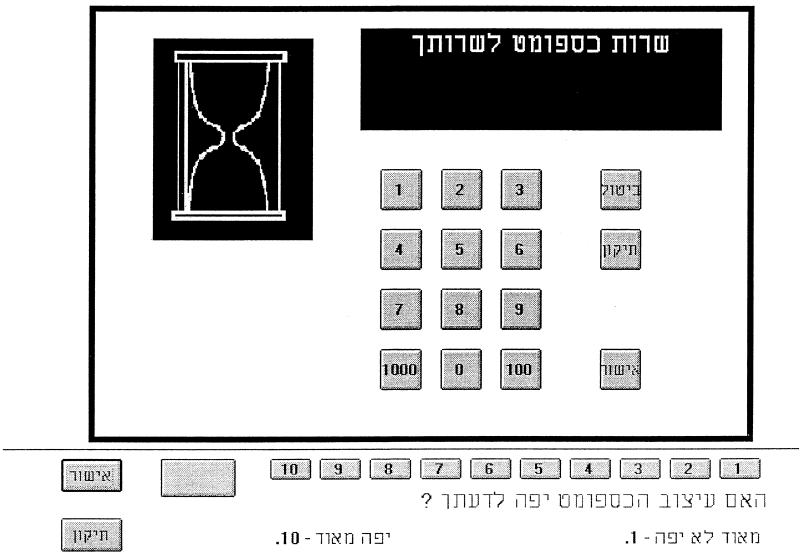
7. Conclusion

This study demonstrated once again the tight relationships between users' initial

perceptions of interface aesthetics and their perceptions of the system's usability. Moreover, we showed that these relations endure even after actual use of the system. We believe that these results shed new light on the role of aesthetics in HCI design and its effects on how users experience their interaction with computerized systems. The results of this study are commensurate with social psychology findings that people associate a person's physical attractiveness with other personal attributes. Similarly, research in the areas of marketing and consumer behavior indicate that aesthetic features of the shopping environment are perceived as related to other, seemingly independent attributes (e.g. functional) of that environment. Obviously, more research is needed to assess the contingencies and boundaries of the aesthetics-usability relationships. Most importantly, these relationships should be studied during a longer time frame than we were able to do. Yet, we believe that there is sufficient evidence already to justify the elevation of the issue of aesthetic design from its current standing at the cellars of HCI research.

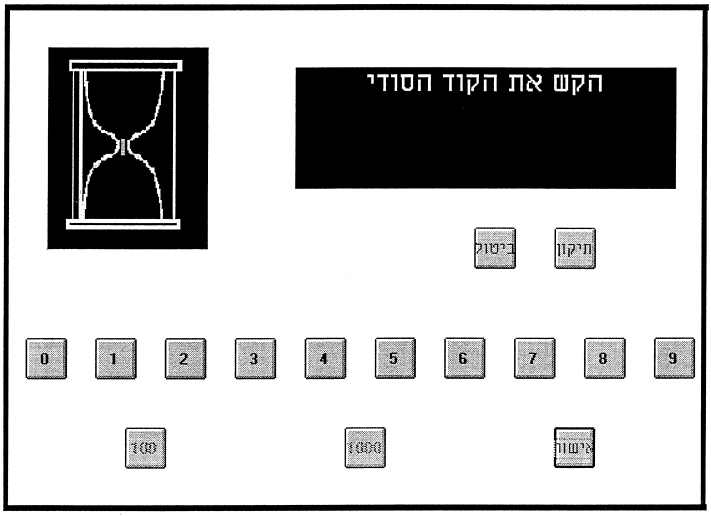
Appendix A

An example of ATM layout rating. The rating question and the rating buttons appear at the bottom of the page below the horizontal bar. In the pre-experimental rating procedure, participants rated this display 8.63 (out of 10) on the aesthetics scale, and 8.39 on the usability scale.

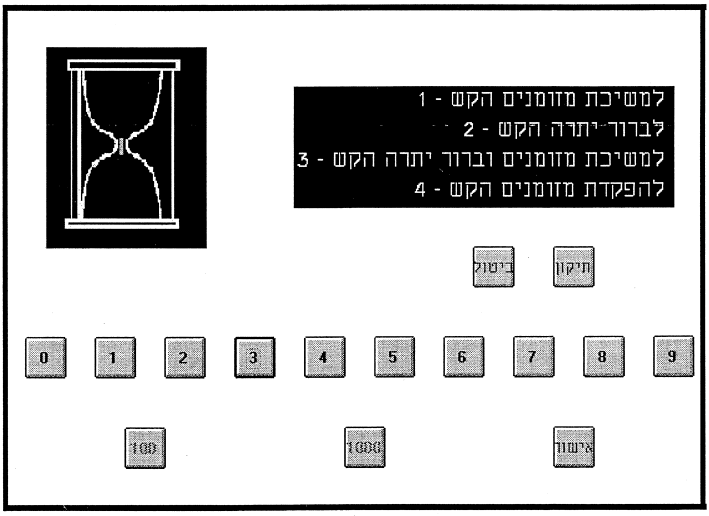


Appendix B

Two sample screens from the task of querying the account balance. In the top screen the user is asked to enter his/her identification number. In the bottom screen the user is asked to select an item from the ATM's main menu. The task is presented to the user at the bottom of the screen below the ATM display. In the pre-experimental rating procedure, participants rated this display 4.10 on the aesthetics scale, and 3.49 on the usability scale.



המטלה: עליך לבצע ברור יתרה.



המטלה: עליך לבצע ברור יתרה.

References

[1] Read, H., 1953. Art and Industry, 3rd ed. Faber and Faber, London.
[2] Lucie-Smith, E., 1983. A History of Industrial Design, Phaidon, Oxford.
[3] Petroski, H., 1993. The Evolution of Useful Things, Knopf, New York.
[4] Norman, D.A., 1988. The Psychology of Everyday Things, Basic Books, New York.
[5] Norman, D.A., 1992. Turn Signals are the Facial Expressions of Automobiles, Addison-Wesley, Reading, MA.
[6] Tufte, E., 1983. The Visual Display of Quantitative Information, Graphics Press.
[7] Alben, L., 1996. Quality of experience: defining the criteria for effective interaction design. Interactions 3 (3), 11–15.
[8] Butler, K.A., 1996. Usability engineering turns 10. Interactions 3 (1), 59–75.
[9] Darden, W.R., Babin, B.J., 1994. Exploring the concept of affective quality: expanding the concept of retail personality. Journal of Business Research 29, 101–109.
[10] Jordan, P.W., 1998. Human factors for pleasure in product use. Applied Ergonomics 29 (1), 25–33.
[11] Kurosu, M., Kashimura, K., 1995. Apparent usability vs. inherent usability, CHI. '95, pp. 292–293.
[12] Tractinsky, N., 1997. Aesthetics and apparent usability: empirically assessing cultural and methodological issues. CHI 97 Conference Proceedings, Atlanta, 22–27 March, ACM, New York, pp. 115–122.
[13] Gould, J.D., 1988. How to design usable systems. In: Helander, M. (Ed.). Handbook of Human–Computer Interaction, Elsevier, Amsterdam.
[14] Shackel, B., 1991. Usability—context, framework, definition, design and evaluation. In: Shackel, B., Richardson, S. (Eds.). Human Factors for Informatics Usability, Cambridge University Press, Cambridge.
[15] Whiteside, J., Bennett, J., Holtzblatt, K., 1988. Usability engineering: our experience and evolution. In: Helander, M. (Ed.). Handbook of Human—Computer Interaction, Elsevier, Amsterdam.
[16] Nielsen, J., 1993. Usability Engineering. AP Professional.
[17] Nielsen, J., Mack, R.L., 1994. Usability Inspection Methods, Wiley, New York.
[18] Shneiderman, B., 1998. Designing the User Interface, 3rd Ed. Addison-Wesley, Reading, MA.
[19] Dix, A., Finlay, J., Abowd, G., Beale, R., 1998. Human–Computer Interaction, 2nd ed. Prentice Hall, Englewood Cliffs, NJ.
[20] Foley, J.D., van Dam, A., Feiner, S.K., Hughes, J.F., 1990. Computer Graphics: Principles and Practice, 2nd ed. Addison-Wesley, Reading, MA.
[21] Marcus, A., 1992. Graphic Design for Electronic Documents and User Interfaces, ACM Press, New York.
[22] Microsoft, 1995. The Windows Interface Guidelines for Software Design, Microsoft Press, Redmond, WA, USA.
[23] Spool, J.M., Scanlon, T., Schroeder, W., Snyder, C., DeAngelo, T., 1999. Web Site Usability: A designer's Guide, Morgan Kaufman, Los Altos, CA.
[24] Laurel, B., 1991. Computers as Theatre, Addison-Wesley, Reading, MA.
[25] Reeves, B., Nass, C., 1996. The Media Equation, CSLI Publications.
[26] Dion, K., Berscheid, E., Walster, E., 1972. What is beautiful is good. Journal of Personality and Social Psychology 24 (3), 285–290.
[27] Eagly, A.H., Ashmore, R.D., Makhijani, M.G., Longo, L.C., 1991. What is beautiful is good, but…: a meta-analytic review of research on the physical attractiveness stereotype. Psychological Bulletin 110 (1), 109–128.
[28] Martineau, P., 1958. The personality of the retail store. Harvard Business Review 36, 47–55.
[29] Russell, J.A., Pratt, G., 1980. A description of the affective quality attributed to environments. Journal of Personality and Social Psychology 38 (2), 311–322.
[30] Donovan, R.J., Rossiter, J.R., 1982. Store atmosphere: an experimental psychology approach. Journal of Retailing 58, 34–57.
[31] Bloch, P., 1995. Seeking the ideal form: product design and consumer response. Journal of Marketing 59, 16–29.
[32] Edell, J.A., Burke, M.C., 1987. The power of feelings in understanding advertising effects. Journal of Consumer Research 14, 421–433.

[33] Derbaix, C., Pham, M.T., 1998. For the development of measures of emotion in marketing: summary and prerequisites. In: Lambkin, M. (Ed.). European Perspectives on Consumer Behaviour, Prentice Hall, Englewood Cliffs, NJ.
[34] Fazio, Roskos-Ewoldsen, Powell, 1994. Attitudes, perception, and attention. In: Niedenthal, P.M., Kitayama, S. (Eds.). The Heart's Eye, Academic Press, New York.
[35] Ross, L., Lepper, M.R., Hubbard, M., 1975. Perseverance in self-perception and social perception: biased attributional process in the Debriefing paradigm. Journal of Personality and Social Psychology 32 (5), 880–892.
[36] Gilbert, D.T., Krull, D.S., Malone, P.S., 1990. Unbelieving the unbelievable: some problems in the rejection of false information. Journal of Personality and Social Psychology 59 (4), 601–613.
[37] Hiltz, S.R., Johnson, K., 1990. User satisfaction with computer mediated communication systems. Management Science 30 (6), 739–764.
[38] Szajna, B., Scamell, R.W., 1993. The effects of information system user expectations on their performance and perceptions. MIS Quarterly 17 (4), 493–525.
[39] Gaeth, G.J., Levin, I.P., Sood, S., Juang, C., Castellucci, J., 1997. Consumers' attitude change across sequences of successful and unsuccessful product usage. Marketing Letters 8 (1), 41–53.
[40] Hays, W.L., 1989. Statistics, 4th ed. Holt, Rinehart and Winston, New York.
[41] Tabachnick, B.G., Fidell, L.S., 1996. Using Multivariate Statistics, 3rd ed. Harper Collins.
[42] Stevens, J., 1986. Applied Multivariate Statistics for the Social Sciences, Lawrence Erlbaum, Hillsdale, NJ.
[43] Zajonc, R.B., 1968. The attitudinal effects of mere exposure. Journal of Personality and Social Psychology 9 (2), 1–27.
[44] Judd, C.M., Brauer, M., 1995. Repitition and evaluation extremity. In: Petty, R.E., Krosnick, J.A. (Eds.). Attitude Strength: Antecedents and Consequences, Lawrence Erlbaum, London.
[45] Hooper, K., 1986. Architectural design: an analogy. In: Norman, D.A., Draper, S.W. (Eds.). User Centered System Design, Lawrence Erlbaum, London.
[46] Isen, A.M., 1983. The influence of positive affect on cognitive organization: some implications for consumer decision making in response to advertising. In: Mitchell, A.A. (Ed.). Advertising Exposure, Memory, and Choice, Lawrence Erlbaum, London.
[47] Berlyne, D.E., 1974. Studies in the New Experimental Aesthetics, Hemispher, Washington, DC.
[48] Martindale, C., Moore, K., Borkum, J., 1990. Aesthetic preference: anomalous findings for Berlyne's psychobiological theory. American Journal of Psychology 103 (1), 53–80.
[49] Veryzer, R.W., Hutchinson, J.W., 1998. The influence of unity and prototypicality on aesthetic responses to new product designs. Journal of Consumer Research 24, 374–394.
[50] Wanous, J.P., Reichers, A.E., Hudy, M.J., 1997. Overall job satisfaction: how good are single-item measures? Journal of Applied Psychology 82 (2), 247–252.