JOURNAL OF APPLIED QUANTITATIVE METHODS

International Symposium on Stochastic Models in Reliability Engineering, Life Sciences and Operations Management (SMRLO'10)

AESTETICS, USEFULNESS AND PERFORMANCE IN USER-SEARCH – ENGINE INTERACTION[1]

Adi KATZ

PhD, Industrial Engineering and Management Department,
SCE - Shamoon College of Engineering,
Beer-Sheva, Israel

E-mail: adis@sce.ac.il

Abstract: *Issues of visual appeal have become an integral part of designing interactive systems. Interface aesthetics may form users' attitudes towards computer applications and information technology. Aesthetics can affect user satisfaction, and influence their willingness to buy or adopt a system. This study follows previous studies that found that users associate aesthetics with other system attributes, e.g. usability. In this study, we asked whether the well-known phenomenon that beautiful things are perceived as good applies to the perception of the system's usefulness. A controlled laboratory experiment tested the relationships between users' perception of aesthetics, usefulness and user performance in tasks performed by participant using an interactive application that surrogated a search engine. We measured users' perceptions of the search engine before and after they used the system to solve information-seeking tasks, and measured user task performance. As expected, significant correlations were found between perceived aesthetics and perceptions of usability and usefulness prior to actual use of the system. We did not find a relation between perceived aesthetics and usefulness after use; and we did not find an expected effect for aesthetic perceptions neither on perceived usefulness nor on performance. We conclude that there is need for a deeper understanding of aesthetic perceptions; a finer grain perspective of perceived aesthetics that differentiates between aesthetic dimensions may reveal that some aesthetic aspects have greater influence on the relations between aesthetics and usefulness.*

Key words: *usefulness; aesthetics; usability; search engines; human computer interactions; interface design*

1. Introduction

The tension between *function* and *form* has long been at the crossroad of artifact design. While emphasis on function stresses the importance of the artifact's usability and usefulness, accentuating the artifact's form serves more the aesthetic and perhaps the social and emotional needs of designers and users (Tractinsky et al., 2000). Today, more and more researchers and interface designers place emphasis on aspects such as aesthetics or promotion of pleasure, and are involved with seeking for opportunities for positive

experiences like pleasure, fun and excitement. There is a remarkable interest in human computer interactions (HCI) to design positive experiences for the user. Designing a good user experience is important not only when designing systems for play and leisure but also for systems that we use for achieving tasks with a well-defined goal. Search engines are a type of such systems. A search engine is an information retrieval system designed to help find information. Its environment enables us to test the relationships between aesthetics and usefulness, because searching for information is considered a task with a well-defined goal that involves decision-making and cognitive effort.

It is important to design positive experiences with systems that we use, because the emotional system changes the way in which the cognitive system operates: emotions change the way the human mind solves problems, and aesthetics can change our emotional state (Norman, 2004). Aesthetics may form user's attitudes towards the system, may improve (or worsen) their performances, affect their satisfaction, and influence their willingness to buy or adopt the system (Tractinsky, 2004).

The main goal of this study is to test whether previously found relations between perceived aesthetics and usability reflect a more general tendency to associate aesthetics with other system attributes. This study focuses on the potential relations between perceived aesthetics and perceived usefulness. In addition, we test whether aesthetics affect performance and user satisfaction. The context of this study is users interacting with a search engine.

The rest of this paper is structured as follows: In the Theory section, we summarize previous studies related to aesthetics of interactive systems and present our propositions. We then refer to usefulness dimensions that are relevant when users evaluate their interaction with search engines. The Method section describes the experimental participants; the apparatus that we designed for the experiment; the experimental design; manipulations, tasks, procedure and the dependent variables' measurements. In the following section, we present and discuss our results and findings. The last section raises the limitations of the current study, its conclusions, and proposes ideas for future work.

2. Theory

2.1. Aesthetics and Positive Experiences in HCI

MIS and HCI have traditionally ignored matters of aesthetics, and whenever aesthetic issues were discussed in the literature and in HCI textbooks, designers were warned against its potential detrimental effects on performance, comprehension, attention and other task-oriented aspects of the interaction. In that perspective, Skog et al. treated aesthetics and utility in as two conflicting concerns that must be reconciled for creating truly useful ambient information visualizations: Visualizations must strike a balance between aesthetical appeal and usefulness (Skog, et al., 2003). Floris claimed that one has to be aware of the possible opposition of utility and attractiveness. There is need for a sensible choice to be made for the relative strengths of the information bearing and the aesthetic factors - including a 'strength zero' of the latter, if need be (Floris, 2008). Lavie and

Tractinsky (2004) discuss the marginalization treatment that the aesthetics dimension receives in the human-computer interaction literature.

The claim in the mid nineties, however, was that modern design places too much emphasis on aspects of performance but not enough emphasis on aspects such as aesthetics or promotion of pleasure. Lavie & Tractinsky claim that any random perusal of websites would suggest that aesthetic considerations are paramount in designing for the web, and report that a new aesthetic wave is increasingly considering aesthetic aspects in human computer interaction. Issues of visual appeal and aesthetics have become an integral part of interactive system design and of information technology (Lavie & Tractinsky, 2004). According to Tractinsky, aesthetics satisfies basic human needs and aesthetic considerations are becoming increasingly important in our society. Today, more and more research and practical design are involved with seeking for opportunities for positive experiences like pleasure, fun and excitement (Tractinsky, 2004). There is a growing recognition of the role of emotional design of everyday things (Norman, 2004) and of information technology systems. IT users need human computer interactions that are complete and satisfying; they deserve an experience that not only achieves task-oriented goals (like efficiency and effectiveness) but also involves the senses and generates positive affective responses (Venkatesh & Brown, 2001).

User experience (UX), a relatively new realm of research in human computer interface design, emphasizes the users' overall satisfaction and experience with a product or a system. While past activities pretty much focused on avoiding negative experiences and on ways in which information technology should be designed to meet user needs for better task performance in terms of efficiency and effectiveness, designing good experiences for users now occupies the HCI community. As the functionality of new information technology products exceed user's needs, and as the prices of systems decrease, the differentiation between products are in terms of UX enhancing rather than on improving functionality (Norman, 1998). One of the various ways to enhance UX is emphasizing aesthetics.

2.1.1. The Positive Effects of Aesthetics

Recently, findings and theories indicate that human decision-making does not rely only on cognitive processes, but also on the affective state (Tractinsky, 2004; Norman, 2004). The emotional system changes how the cognitive system operates: emotions change the way the human mind solves problems, and aesthetics can change our emotional state (Norman, 2004). Affect changes how well we do cognitive tasks: affect regulates how we solve problems and perform. Negative affect can make it harder to do even easy tasks, while positive affect can make it easier to do difficult tasks (Norman, 2002). Following this idea, this research will test whether aesthetic interfaces affect performance in the context of users interacting with a search engine.

Proposition 1: *Users' aesthetic perceptions of a system have an affect on their performance in the system: users perform better with search engine that they perceive as more beautiful.*

In pleasant, positive situations, people are much more likely to be tolerant of minor difficulties and irrelevancies. Although poor design is never excusable, when people are in a relaxed situation, the pleasant, pleasurable aspects of the design will make them more tolerant of difficulties and problems in the interface (Norman, 2002). Following this logic, we

expected that users would perceive beautiful search engines as useful and satisfying, even when the usefulness of the search engine is low.

Proposition 2: *Users' aesthetic perceptions of a system have an affect on their satisfaction with the system: users are more satisfied with search engine that they perceive as more beautiful.*

2.2. Usefulness, Usability and Aesthetics

Usefulness is the issue of whether the system can be used to achieve some desired goal. *Perceived Usefulness* is the degree to which a person believes that using a particular system would enhance his or her job performance (Davis, 1989, p. 320). People form perceived usefulness judgments in part by cognitively comparing what a system is capable of doing with what they need to be done by their job. TAM2 (The extended technology acceptance model that models how users come to accept and use a technology) theorizes that people use a mental representation for assessing the match between job goals and the consequences of performing the act of using a system as a basis for forming judgments about the use-performance contingency. One key component of the matching process is the user's cognitive judgment of job relevance that exerts a direct effect on perceived usefulness (Venkatesh and Davis, 2000).

Usability is a quality attribute that assesses how easy a user interfaces is to use, and is defined by five quality components: learnability, efficiency, memorability, errors and satisfaction (Nielsen, 1993). *Perceived Usability* is the degree to which a person believes that using a particular system would be free of physical and mental effort (Davis, 1993, p. 477).

2.2.1. Relations of Aesthetics with Usability and Usefulness

It was found that aesthetics are highly correlated with perceptions of the system's usability before (Tractinsky, 1997) and after (Tractinsky et al, 2000) the interaction. Aesthetics may form user's attitudes towards the system, may improve (or worsen) their performances, affect their satisfaction, and influence their willingness to buy or adopt the system (Tractinsky, 2004). Aesthetic impressions may affect how people perceive other attributes of a system, like usability or ease of use (Tractinsky et al., 2000) and perceived goodness of a system (Hassenzahl, 2004). Lavie and Tractinsky (2004) found a relationship between the aesthetics factor and the perceived service quality of a web site, and say that it is possible that aesthetics is the primal factor affecting other perceptions.

Proposition 3: *Aesthetic perceptions of systems are related to usability perceptions. A search engines that is perceived as beautiful is also perceived as usable.*

In the ancient world, judgments of a product's usefulness and beauty were one of the same (Lavie and Tractinsky, 2004). However, correlations between aesthetics and perceived usefulness had not been investigated experimentally. One of the goals of this study is to test whether previously found correlations between perceived aesthetics and usability reflect a more general tendency to associate aesthetics with other system attributes. Perhaps a halo effect may cause carry over of an aesthetic design to perceptions of other design features (Tractinsky et al., 2000). We focus on the potential relation between

perceived aesthetics and perceived usefulness and wish to find whether the well-known phenomenon that beautiful things are perceived as good applies to the perception of system's usefulness.

Proposition 4: *Aesthetic perceptions of systems are related to usefulness perceptions. A search engines that is perceived as beautiful is also perceived as useful.*

2.3. Usefulness of Search Engines

A search engine should allow users to compose their own search queries rather than simply follow pre-specified search paths or hierarchy as in the case of certain catalogues (Chu & Rosenthal, 1996).

2.3.1. Content Relevancy

The main purpose of a search engines is to retrieve the relevant documents for a given request. It is therefore natural that the literature on search engines and retrieval systems has to a large degree concentrated on relevance-oriented questions, i.e., on the relevance and precision of the retrieved results. *Content relevance* is defined as the adequacy of the content of a document in response to the request. Subjective relevance is defined as the usefulness of the document to the user (Bing & Harvold, 1977).

2.3.2. Informative Results

In addition to content relevancy, another aspect of a search engine's usefulness is the degree to which search results are informative. A SERP (Search Engine Results Page) listing contains a list of links to web pages along with a short summary of the pages. Those pages include content that matches the search terms. The usefulness of a web page's description varies on the extent that it conveys helpful information, ranging from descriptions that reveal the answer to the research question (most informative) through descriptions that reveal the content of the web site they represent (informative), to descriptions appearing in gibberish (uninformative). According to Kowalski, there is a likely possibility that there will be items found by the query that are not retrieved by the user for review (Kowalski, 1997). Users will not review items in the SERP listing when the summary of information in the display is sufficient to judge that the item is irrelevant. Usefulness is higher on one hand whenever users are able to avoid accessing into fruitless pages, and on the other hand, when they are able to access into useful pages on the outset. Informative results are in line with Lancaster and Fayen's (1973) form of output dimension that refers to the various formats in which the documents and feedback indicators may be presented to the user and with TAM2's output quality notion, a determinant of perceived usefulness (Venkatesh and Davis, 2000).

In the following section, we describe in detail a laboratory experiment that we conducted to test the propositions. We will also refer to content relevancy and to informative results when we describe the usefulness manipulation.

3. Method

The above propositions were tested in a laboratory experiment. Participants interacted with a computer application that served as a surrogate for a search engine, to find answers for search tasks. Two variables were manipulated: *aesthetics* of the search engine's screen layout, and *usefulness* of the search results. We manipulated aesthetics by allocating subjects to work with a system at a certain level of aesthetics, based on their prior evaluation of the beauty of different screen layouts. Usefulness was manipulated based on two dimensions: the relevancy of the results to the question in task, and the brief summary information conveyed by the site's link in the SERP listings. Below we describe the experimental participants; the participants; the apparatus that we designed for the experiment to surrogated a search engine; the experimental design; the manipulations, tasks, procedure, and the dependent variables' measurements.

3.1. Participants

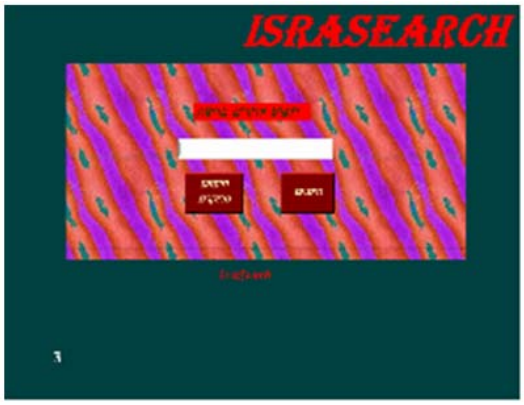
Sixty Israeli undergraduate students from a College of Engineering participated in the experiment, all of them in their third year, and all of them specializing in Information Systems. They received class credit for their participation as part of their "Human Computer Interactions" course. In addition, they were aware of the possibility that the three top performers in the experiment might receive monetary prize. There were 47 males and 17 females, and their ages ranged from 20 to 34 years, with an average age of 26.68. Sixty seven percent of them use search engines frequently (very often or every day) and the rest are familiar with search engines but use them only occasionally.

3.2. Apparatus

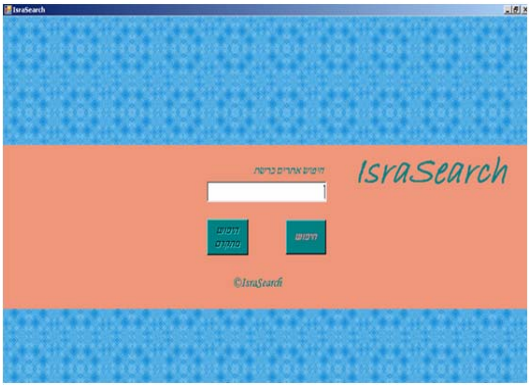
For the research, we built a computer application named IsraSearch, surrogating a search engine. The reason that we did not use a real engine was to ensure experimental control over certain variables that we did not manipulate but might bring potential noise to the experiment (such as the number of links in the SERP listings).

Figure 1 presents the opening page of six interface layouts that were used in the experiment. Israsearch's appearance imitated a real search engine such as Google. There were six different IsraSearch interfaces, identical in their controls and displayed elements. The only difference between them was their aesthetic, in terms of their "skin", i.e., colors of the elements, background textures, font styles, and locations of two captions. The choice of interfaces was based on a pilot with 30 undergraduate students (who did not participate in the main experiment). For the pilot, we designed 32 interface layouts. Each interface included the basic search controls; a textbox for typing search terms and a "search" button, to appear like common search engines. We presented them to the students on a big screen. Afterwards each student sat in front of a personal computer screen, observed the same designs individually, and rated the interfaces

Low on aesthetics



Medium on aesthetics



High on aesthetics

Figure 1. The six IsraSearch interface designs

on a 5 point Likert scale, from "very unattractive" (1) to "very beautiful" (5). Based on these ratings, we chose six designs for the main experiment, of which two were rated as highly aesthetic, two as low in terms of aesthetics, and the remaining two received medium ratings. In Figure 1, we present the six experimental designs arranged in three rows based on their aesthetic ascription in the pilot: low, medium and high, respectively.

The display of SERP (Search Engine Results Page) listings too resembled the format of other search engines, such as the changing colors of visited links, a headline and a short summary describing the web page to be accessed by each link. An example for a SERP listing is presented in Figure 2. The interface and search language were Hebrew.

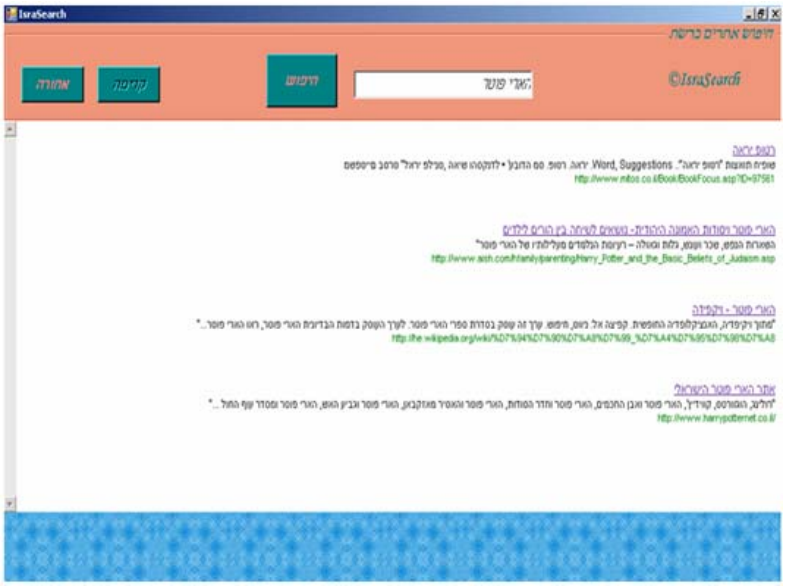


Figure 2. Example of an IsraSearch SERP listing

Being a surrogate for a real search engine, IsraSearch does not have an indexing algorithm. Instead, the pre-selected SERP lists were displayed only if the user's search statement contained at least a minimum set of predefined search terms. The minimum set of predefined search terms was determined for each task in advance, based on the results of a survey we conducted with 10 subjects who did not participate in the experiment. The survey participants were presented with 10 experimental tasks (queries). As each query defines the formal properties that a document must have in order to be retrieved (Bing & Harvold, 1977), we asked them to write a list of appropriate search terms they would type in order to retrieve pages that will help them solve the task. We summed the various terms produces by our participants for each task to a list, and from each list we chose a minimum set of terms that were most frequent (these terms represent the "core" properties that a document must have in order to be retrieved). At the experiment, search results for each query were actually web pages that were selected in advance and were presented to the user whenever he entered the minimum predefined set of search terms. Of course, experimental participants were not informed that the search engine they use is only a surrogate for a real engine.

3.3. Experimental Design

The experiment used a 3 (between) X 3 (within) factorial design. The between groups factor was the aesthetic level of the interface and the within-subjects factor was the usefulness of the search results. Both factors had three levels: low, medium and high. We now explain the manipulations of the aesthetic and the usefulness factor.

3.4. Manipulations

3.4.1. Aesthetics Factor

To obtain three levels of aesthetic interfaces, we randomly allocated subjects to three different experimental conditions. When participants arrived, we first asked each to "blindly" draw a scrap of paper with a system login. Each drawn login was assigned in advance to one of three aesthetic conditions: low, medium or high. Then, each participant sat in front of a personal computer screen, and was exposed to the six IsraSearch interface designs (shown in Figure 2) separately. At each appearance of a design, participants were asked to rate it on a 1-5 Likert scale with regard to three attributes: Aesthetics, Usability and Usefulness. Therefore, each participant made 18 ratings at this phase. We randomized the appearance of designs and rating items. After the rating phase, each subject was assigned to interact with only one of the six screen layouts based on two determinants: 1) the login he raffled at the beginning; and 2) his own ratings of the designs. For example, a participant who drew a login that was assigned to the low aesthetics condition worked with a design that he rated as least aesthetic.

3.4.2. Usefulness Factor

Each participant had to perform 10 search tasks, in each the participant had to answer 10 questions by conducting searches in IsraSearch. To obtain three levels of usefulness, we manipulated two attributes of the search results in an additive way: 1) content relevancy of the results; 2) the degree to which the results were informative. We created a mixture of the two attributes to differentiate between three groups of tasks that ranged from high usefulness to low usefulness. For the 10 searches, the experimental software returned 4 result sets with high usefulness results, 4 result sets with low usefulness results, and 2 sets with medium usefulness results. The ten search tasks are presented in Table 1, grouped in terms of their level of usefulness (high, medium and low). We explain the mixture of content relevancy and the degree to which the results were informative below.

Table 1. The ten search tasks

Usefulness	Task #	Task
high	1	**Which books did Harlen Kuben write?**
	2	**What are isobars?**
	3	**Alexander Mokdon was the son of:**
	4	**What is the height of the first floor at the Eifel tower?**
medium	5	**A poodle dog has several possible sizes. What is the size (height and weight) of the medium poodle?**
	6	**The K1200S engine of a BMW has:**
low	7	**What is the origin of the walnut?**
	8	**Surami is a kind of food that comes from:**
	9	**The turmeric's medical qualities are:**
	10	**The last movie in which Ingrid Bergman played in is:**

3.4.2.1. Content Relevancy

Relevancy has to do with whether the information appearing in a web site is sufficiently specific to answer the user's problem. The manipulation of this dimension is in line with TAM's perceived usefulness construct (Davis, 1989; Venkatesh and Davis, 2000) as previously mentioned, and with previous studies in which users estimate the relevancy of retrieved results (e.g., Shapira et al., 2005; Pan et al. (2007); Coiera and Vickland, 2008).

Content relevancy is defined as the adequacy of the content of a document in response to the request. Users' perception of the content relevancy reflects the usefulness of the document to the user. For a given document in a given situation, users will normally be able to decide whether or not the document is clearly irrelevant, or whether it might be relevant (Bing & Harvold, 1977). Only users can make valid judgments regarding the suitability of information to solve their information need (Kowalski, 1997).

As aforementioned, when a subject entered the minimal set of required search terms, a SERP listing was displayed on the screen. Each SERP listing contained four results (for uniformity and for controlling other sources of variance).

- High usefulness level was achieved by a mixture of search results that had a relatively large proportion of relevant web sites and a relatively low proportion of pages with partial or little relevancy. Highly Relevant pages are ones that contained a full answer to the question asked; partial relevancy pages contained only a portion of the answer, and low relevancy pages are related to the object in question but not to the specific question regarding that object.
- Medium usefulness level was achieved with a mixture of a relatively large proportion of pages with partial relevancy and only a small proportion of relevant or irrelevant pages. Irrelevant pages contained the object referred to in the question but had a very weak relation to that object. In other words, those pages were mainly about other topics and the object in question was only slightly mentioned in them.
- Low usefulness level was achieved by a mixture of web pages that had a relatively large proportion of irrelevant web sites and a relatively low proportion of pages with partial relevancy.

We demonstrate this using task# 5: "A poodle dog has several possible sizes. What is the size (height and weight) of the medium poodle?" A highly relevant page contained a list of all possible poodle sizes; a page with partial relevancy contained information about the height of the medium poodle but no information regarding its weight; a page with low relevancy presented poodles for adoption; and an irrelevant page was an article about a soccer couch with the title "If you want to coach, you need to be a poodle".

3.4.2.2. Web Sites Degree of Informative Results

In addition to content relevancy, the second attribute of usefulness that we manipulated was the degree to which the search results were informative. Resembling real search engines, each SERP listing of IsraSearch contained a list of web pages with titles, a link to each page, and a short summary showing where the search terms have matched content within the page (see Figure 2). The usefulness of a web page's description varied on the extent that it conveyed helpful information, ranging from descriptions that reveal the answer to the research question (highly useful; most informative), through descriptions that contain the content of the web site they represent (medium usefulness; informative), to descriptions appearing in gibberish (low usefulness; uninformative). When tailoring page summaries to achieve different usefulness levels, the mixture of 4 web site links varied in the proportion of informative and uninformative page descriptions: the proportion of informative page descriptions were increased and the proportion of uninformative page description were decreased when advancing from low to high usefulness results. The degree of informative

results was adopted as a usefulness dimension from Kowalski (1997) as described earlier in the theory, when referring to search engine's usefulness.

Table 2 summarizes the manipulation of the two independent variables: aesthetic factor (between) and the usefulness factor (within).

Table 2. Experimental factors (manipulated variables)

Variable	Explanation	Values	Point of Measurement
Aesthetics	Manipulated by assigning each subject to a certain aesthetic condition that matched his system login and his ratings of 6 screen layouts	Low, Medium, High	Pre-experiment
Usefulness	Manipulated by affecting two usefulness dimensions: content relevancy of web sites for each task, and the summary information displayed in the SERP listing.	Low, Medium, High	Pre-experiment

3.5. The Experimental Procedure

Participants worked in computer labs under the experimenters' supervision. Each of them sat separately in front of a personal computer. They first had to type their system login that they blindly drew, as described previously. The experimental session included three stages: 1) layout rating; 2) practice task; and 3) experimental tasks.

1. Layout rating: Participants rated the six IsraSearch interface designs (presented in Figure 1) on three attributes: aesthetics, usability and usefulness. We measured usability perceptions to replicate aforementioned studies that found a relationship between users' perceptions of a system's aesthetics and usability. We used aesthetic ratings before use, to examine its relation with other system attributes perceived by the user (usability and usefulness) and we also used them to manipulate the aesthetic factor as explained earlier in the aesthetics factor section.

2. Practice task: After receiving instructions about the search engine and about the task, they practiced the use of IsraSearch by performing one preliminary search task that we used for practice At this stage, participants already worked with the screen layout that they were assigned to for the experiment (as explained earlier in the aesthetics factor section).

3. Experimental tasks: After the practice stage, participants began the experimental stage, searching the IsraSearch "engine" to answer the 10 questions (see Table 1). The questions were related to various topics, and were deliberately not trivial; that is to say that participants had to use IsraSearch to find full and correct answers. Each question was presented at the bottom of the screen, one at a time, in a random order (different for each participant), along with a set of 4 possible answers, in which only one was correct. To delimit the experiment to a reasonable time range and to raise participant's motivation and arousal, we set a limit of 5 minutes for answering each question, assuming that it is sufficient for finding the answer (in usefulness groups containing the answer in the search results). Each task ended when the subject chose one answer, by clicking on one out of 4 radio buttons and submitting the answer by clicking on a "send answer" button. If the 5-minutes time limit had run out and no answer was chosen (a timer was presented at the bottom of the screen), the task was stopped and the

following task was presented. We informed participants that their goal is to achieve a maximum number of successful answers in a minimum time range.

The experimental task resembles realistic interactions with retrieval systems, in which the user determines the information he needs and creates a search statement. The system processes the search statement, returning potential hits displayed in SERP listings. Resembling real search engines, the summaries of web pages in the SERP listing are descriptions varying from most informative (helpful in revealing the answer to the research question) to uninformative descriptions (appearing in gibberish). Then, the user selects items from the list to review and access. Resembling real search engines, the content of selected web pages varied in relevancy in terms of the adequacy of the content in response to the request.

3.6 Experimental Dependent Measures

We measured five dependent variables: perceptions of usefulness, usability, and aesthetics, user satisfaction and performance. Table 3 summarizes the dependent variables by this order. All dependent measures except for performance used five-point Likert scale items.

At the first stage, previously referred as the *layout rating stage*, Participants rated the six IsraSearch interface designs (presented in Figure 1) on a 1-5 Likert scale with regard to three attributes: aesthetics, usability and usefulness. We measured usability perceptions to replicate aforementioned studies that found a relationship between users' perceptions of a system's aesthetics and usability.

At the *experimental stage*, upon completion of each experimental task, each subject presented with four 5-point Likert-type statements that asked him/her to rate the engine on four attributes: one general usefulness question, two additional usefulness questions reflecting the two usefulness attributes we manipulated (content relevancy and degree of informative results) and the subjective satisfaction with the engine. Each statement was displayed separately, in a window that popped in the middle of the screen. At the end of the experimental stage, upon completion of all experimental tasks, an additional popup window presented again the same four Likert-type statements, this time referring to the search engine generally, that is to say beyond individual tasks.

The system's log recorded various performance measurements during the experimental stage: the number of search iterations, number of visited links (sites), time to complete each task, and the number of successful tasks. Measuring user performance by the *number of search iterations* follows Shapira et al. (2005) who measured user effort by the number of iterations required to perform a task, considering each query submitted as a single iteration. More search iterations for a given task reflects a low usefulness level because the user needs to engage in repeated searches to achieve satisfactory results for accomplishing the task. The same logic of performance efficiency was applied in two additional measures for performance: *number of visited links* and *the time it took to complete each task*. Assuming that more correct answers in a limited time range, exhibit higher performance, user performance was also measured by the overall *number of successful answers*.

4. Results

In this section, we present the results for the experiment. We start with the results of the manipulation checks, and present an examination we conducted to make sure that our participants were able to sense usefulness properly.

4.1. Manipulations Checks

4.1.1. Aesthetics Manipulation Check

As described in the Method section, aesthetics was manipulated by allocating subjects to work with a system at a certain level of aesthetics, based on their prior evaluation of the beauty of different screen layouts. A successful manipulation of aesthetics is one that produces three aesthetic groups whose perceptions of aesthetics are significantly different, each composed of participants that consider the layout they worked with at a compatible aesthetics level. A ddddddone-way analysis of variance (ANOVA) revealed a significant effect of the aesthetic factor: $F(2, 57) = 83.98$, $p < .001$. Mean ratings of IsraSearch's aesthetics were 3.53, SD = 0.125; M = 3.00, SD = 0.108; M = 1.45, SD = 0.115 for high, medium and low aesthetic conditions, respectively. Scheffe post hoc contrasts to test whether the differences between any pair of three conditions were statistically significant revealed a significant difference at 0.001 between the low and the other two conditions, and a difference at the 0.05 between the high and medium conditions. The results indicate that the aesthetics manipulation was successful. Indeed, the high aesthetic group was composed of participants who worked with a design that they ascribed as beautiful, while the medium aesthetic group was composed of participants who worked with a design that they did not consider as beautiful nor as ugly, and the low aesthetic group was composed of participants who worked with a design that they ascribed as ugly.

Table 3. Dependent variables measurements

Variable	Point of Measurement	Item
Usefulness	Rating stage - subjective valuation of usefulness based on the system's screen layout	"What is your evaluation of the system's usefulness?"
	Experimental stage, after each task and after completion of all tasks - three subjective valuation of the system's usefulness	a. "Were the web pages provided by the search engine relevant to the task?" (Relevancy dimension) b. "Was the information conveyed by the page summaries in the SERP listing helpful?" (informative results dimension) c. "Were the search results appropriate?" (General usefulness)
Usability	Rating stage - subjective valuation of the system's usability based on its screen layout	"How easy is it to use the search engine?"
Aesthetics	Rating stage - subjective valuation of the screen layout's aesthetics	"What is your evaluation of the system's aesthetics?"
User Satisfaction	Experimental stage, after each task	"Are you satisfied with the search results for this task?"
	After completion of all tasks	"Are you satisfied with the search results for the various tasks?"
User Performance	Experimental stage, after each task - objective measurements of user's achievement in each task	a. Correctness of the chosen answer (true/false). b. Number of search iterations per task, that is- the number of times a subject stimulates a search by clicking on the search button (after entering search terms). c. Number of visited links (the number of times a subject clicked on the links appearing in the SERP listing d. Time to complete each task

Variable	Point of Measurement	Item
	After completion of all tasks - objective measurements of user's overall achievement	a. Number of successful answers (maximum of 10 correct) b. Overall and average number of search iterations) c. Overall and average number of SERP links clicked d. Overall time to complete all tasks and average time to complete tasks

4.1.2. Usefulness Manipulation Check

As described earlier, we manipulated usefulness by creating for each search task a certain combination of results' relevancy in the SERP listing, and a certain degree for which the results were informative. Four tasks were characterized by result sets with high usefulness, other four were characterized by result sets with low usefulness, and two additional tasks had results sets with medium usefulness.

Using repeated measures, we tested the difference in performance between the three different levels of task usefulness in terms of time to complete the task, the number of search iterations per task, the percentage of correct answers, and the number of visited links. The results are presented in Table 4.

Table 4. Repeated measures for difference in performance between three usefulness groups

Measure	Within Subjects Effect	Means and STD		
		High usefulness	Medium usefulness	Low usefulness
Time (seconds)	F(2,58) =7.16, p < .001	85.51 (6.35)	121.45 (11. 41)	169.97 (8.84)
Number of Search iterations	F(2,58) =28.20, p < .001	7.02 (6.12)	8.27 (8.50)	12.47 (9.55)
Success (% of correct answers)	F(2,58) =130.45, p < .001	87.5% (0.20)	38.3% (0.23)	42.92% (0.21)
Number of visited links	F(2,58) =256.42, p < .001	6.22 (5.50)	13.24 (7.34)	16.15 (7.63)

For the time to complete the task and for the number of search iterations, the results are consistent with the expected pattern for each usefulness level revealing a significant within-subjects effect. For percentage of correct answers, and for the number of visited links, a significant difference was found only between high usefulness tasks and the low and medium usefulness levels. However, we found no significant difference between the medium and the low usefulness levels. This means that the usefulness manipulation did not significantly differentiate between the medium and the low usefulness levels. To deal with this finding, we checked whether certain tasks were misplaced by closely examining performance measures for each task. A following and separate examination of the performance measures for each task in IsraSearch's log revealed that Task 6 (see Table 1), ascribed as medium usefulness, was problematic, having an exceptional amount of search iterations. The fact that this was the only task that required typing keyword in a combination of English and Hebrew, might explain this. In addition, Task 5, ascribed as medium on usefulness, was actually easy, as it took very little time to complete, required only a few search iterations, and only one subject failed to find the right answer. Table 5 presents the means and the standard deviations of time to complete each task (in seconds), number of search iterations, and percentage of success for each task. Based on these results, we decided to omit Task 6, and ascribed Task 5 to the high usefulness task group. A following manipulation check referred to the two remaining usefulness groups: high versus low.

We continued with repeated measures analysis to examine the difference in performance between the two remaining usefulness levels, in terms of time to complete the task, number of search iterations per task, percentage of correct answers, and number of visited links. The results are presented in Table 6.

The results in Table 6 show that subjects perform better when usefulness is higher. Therefore, the new ascription of tasks to two levels of usefulness following the manipulation check is effective.

Table 5. Performance measures for each search task

Useful-ness	Question/Task (abbreviated titles)	Time		Search iterations		Success	
		Means	STD. Dev	Means	STD. Dev	% Correct	STD. Dev
High	1) Harlen Kuben	85.18	75.09	1.57	1.94	90	0.30
	2) Isobars	105.22	143.65	1.17	0.64	81.67	0.39
	3) Mokdon	69.10	49.31	2.77	5.61	83.33	0.38
	4) Eifel tower	82.55	51.32	1.52	1.64	95	0.22
Medium	5) Poodle dog	100.87	163.07	1.28	0.99	98.33	0.13
	6) BMW Engine	142.03	72.01	6.98	8.28	75	0.44
Low	7) Walnut	157.85	63.88	2.18	2.04	66.7	0.25
	8) Surami	187.10	80.50	3.22	2.49	30	0.46
	9) Turmeric	161.53	77.50	3.25	3.05	73.33	0.45
	10) Ingrid Bergman	173.38	160.46	3.82	5.57	61.67	0.49

Table 6. Repeated measures for difference in performance between the two remaining usefulness groups

Measure	Within Subjects Effect	Means and STD	
		High usefulness	Low usefulness
Time (seconds)	F(1,59) =75.08, p < .001	88.58 (52.74)	169.97 (68.50)
Number of search iterations	F(1,59) =75.11, p < .001	1.66 (1.24)	12.47 (9.55)
Success (% of correct answers)	F(1,59) =98.10, p < .001	70.33 (0.16)	42.92 (0.21)
Number of visited links	F(1,59) =245.82, p < .001	1.15 (0.50)	16.15 (7.63)

4.2. Verification of Participants' Usefulness Perceptions

We previously claimed that users are normally able to decide whether or not a document is clearly irrelevant, or whether it might be relevant (Bing & Harvold, 1977), and that only users can make valid judgments regarding the suitability of information to solve their information need (Kowalski, 1997). Therefore, we examined our participants' ability to have a good sense of usefulness, by looking at their usefulness ratings for each search task. As described earlier, after each task, four questions appeared in a pop-up window, referring to that task. The first three questions were about usefulness, while the fourth was about their satisfaction with the search results.

Results of a repeated-measures ANOVA for each usefulness question that popped-up, revealed a significant overall difference between tasks characterized by low versus high usefulness. Table 7 presents repeated measures for the difference in perceived usefulness and satisfaction between usefulness groups.

Table 7. Repeated measures for difference in perceived usefulness and satisfaction between usefulness groups

Measure		Within Subjects Effect	Means and STD	
			High usefulness	Low usefulness
Usefulness	Relevancy dimension	F(1,59) =237.53 p < .001	4.07 (0.65)	2.57 (0.73)
	Informative results dimension	F (1, 59) =221.29 p < .001	3.93 (0.67)	2.56 (0.80)
	Usefulness - general item	F (1, 59) =173.12 p < .001	4.09 (0.65)	2.61 (0.76)
User satisfaction		F (1, 59) =164.88 p < .001	4.16 (0.68)	2.56 (0.81)

For the first question, "Were the web pages provided by the search engine relevant to the task?" (Relevancy dimension) high usefulness tasks were rated as having significantly more relevant web pages than low usefulness tasks. Participants were able to sense usefulness properly, in other words, they were able to distinguish between results characterized by a high relevancy of content and results that were low on content relevancy.

For the question "Was the information conveyed by the page summaries in the SERP listing helpful?" (Informative results dimension), high usefulness tasks were rated as having significantly more helpful SERP listing than low usefulness tasks. Participants were able to sense usefulness in terms of the usefulness of page summaries in the SERP listings.

For the question, "Were the search results appropriate?" (a general usefulness-item), high usefulness tasks were rated as having significantly more appropriate results than low usefulness tasks. Participants were able to sense the appropriateness of the results to the task they were conducting.

Results of a repeated-measure ANOVA for satisfaction revealed a significant overall difference between tasks characterized by low versus high usefulness. For the forth question, "Are you satisfied with the search results for this task?", high usefulness tasks results were rated as more satisfying than low usefulness tasks results. Participants were satisfies when search results were informative and relevant to their task.

4.3. Propositions 1-2

4.3.1. Proposition 1

To test Proposition 1, that aesthetics of search engines will affect performance, a test of a between-subject affect for aesthetics on all performance measures was conducted using MANOVA. Table 8 shows descriptive statistics for each performance measure and results of F-tests of the aesthetic effect. For all performance measures, there was no effect of the aesthetics factor. The user's aesthetic perception of a search engine did not affect his performance in the task of information searching.

Table 8. Performance measurements in each aesthetic group

Performance measure	Aesthetic group*	Mean and Std.	F (df)
Overall number of successful answers	High	5.88 (1.22)	
	Medium	6.22 (1.62)	0.466 (2,57)
	Low	5.80 (1.58)	
Overall time to complete all tasks (m-sec)	High	1185.94 (343.44)	
	Medium	1223.96 (334.23)	0.197 (2,57)
	Low	1153.95 (416.71)	
Average time to complete tasks (m-sec)	High	127.31 (51.67)	
	Medium	135.06 (47.72)	0.905 (2,57)
	Low	115.91 (40.59)	
Overall number of search iterations	High	27.23 (15.18)	
	Medium	26.69 (14.56)	0.213 (2,57)
	Low	29.40 (12.50)	
Average number of search iterations	High	2.72 (1.52)	
	Medium	2.67 (1.46)	0.213 (2,57)
	Low	2.94 (1.25)	
Overall number of SERP links clicked	High	23.29 (9.07)	
	Medium	24.48 (7.90)	0.789 (2,57)
	Low	20.95 (10.79)	
Average number of SERP links clicked	High	2.33 (0.91)	
	Medium	2.45 (0.79)	0.789 (2,57)
	Low	2.09 (1.08)	

**Number of subjects is 17, 23 and 20, for high, medium and low aesthetic groups, respectively*

4.3.2. Proposition 2

To test Proposition 2, that aesthetics of search engines will affect user satisfaction, we tested the between-subject affect for aesthetics on post-satisfaction using ANOVA. Table 9 shows that we did not find an effect for the aesthetics factor on satisfaction as expected in Proposition 2. In other words, user's aesthetic perception of a search engine did not affect the degree of satisfaction with it.

Table 9. Post-satisfaction in each aesthetic group

Aesthetic group	Mean and Std.	F (df)
High	3.35, SD = 0. 20	
Medium	3.59, SD = 0.18	1.911 (2,56)
Low	3.10, SD = 0.17	

4.4. Propositions 3-4 and Additional Intercorrelations

4.4.1. Proposition 3

We measured perceptions of the search engine before (layout rating stage), and after (popup questions at the end of the experiment) the actual use of IsraSearch. We measured perceived usefulness, perceived usability, perceived aesthetics, and user satisfaction. Intercorrelations among the perceived aspects both before and after the experiment are presented in Table 10.

Table 10: Pearson correlation matrix of pre- and post experimental perceived measures

	Pre-Aesthetics	Pre-Usefulness	Pre-Usability	Post-Usefulness1	Post-Usefulness2	Post-Usefulness3	Post-Satisfaction
Pre-Aesthetics	-	0.610**	0.402**	0.093	0.063	0.145	0.243*
Pre-Usefulness		-	0.432**	0.203	0.150	0.108	0.205
Pre-Usability			-	0.142	0.183	-0.103	0.005
Post-Usefulness1				-	0.634**	0.523**	0.710**
Post-Usefulness2					-	0.527**	0.537**
Post-Usefulness3						-	0.598**
Post-Satisfaction							-

*** Correlation is significant at the 0.01 level (1-tailed), N=60*
Post-usefulness dimensions: 1 - relevancy; 2 - informative results; 3 - general usefulness

As proposition-3 predicted, pre-use perceptions of IsraSearch's aesthetics and their perceived usability are significantly correlated. This follows previous studies by e.g. Kurosu & Kashimora (1995) and Tractinsky et al. (2000), who found that users associate aesthetics with usability.

4.4.2. Proposition 4

Pre-use perceptions of IsraSearch's aesthetics and their perceived usefulness are significantly correlated as expected in proposition 4, and are high as the correlations between aesthetics and perceived usability obtained by Kurosu & Kashimora (1995) and Tractinsky et al. (2000).

While correlation between pre-experimental perception of aesthetics and post-experimental perceived usability were significant (0.5) in the study of Tractinsky et al. (2000), correlations between pre-experimental perceived aesthetics and post-usefulness perceptions were diminished at the end of the experiment (as shown in table 10). System layouts that were considered as aesthetic before use were not perceived as more useful after using the system. We will refer to this result in the study limitation section.

As expected in Propositions 3-4, pre-use perceptions of IsraSearch's aesthetics and pre perceptions of usability and usefulness are significantly correlated, suggesting that a halo effect causes carry over of aesthetics on other perceptions of a search engine. Prior to actual use, search engines that are perceived as beautiful are also perceived as more usable and as more useful.

4.4.3. Additional Correlations

We originally formulated two propositions (3-4) which refer to correlations between aesthetic perceptions and other users' perceptions of the search engine, and found additional correlations in our results. Table 10 shows that pre perceptions of usability and usefulness are significantly correlated. This relation was not central in the current research, but is not surprising; TAM (Technology Acceptance Theory) argues that perceived usefulness is influenced by perceived ease of use. The easier a system is to use, the more useful it can be (Venkatesh and Davis, 2000). The three post-use perceptions of usefulness are significantly inter-correlated. In addition, they are all correlated with overall satisfaction with the system with remarkably high correlation between the relevancy dimension and satisfaction. Search engines that return highly relevant web pages satisfy their users. An interesting finding is that while post-use satisfaction is uncorrelated with pre-use perceptions of usability and usefulness, it is significantly correlated with pre-use perception of aesthetics. Users were satisfied with search engines with layouts that were aesthetically pleasing but

were not necessarily satisfied with search engines with a layout that seemed usable or useful for searching information before use.

5. Discussion and Conclusions

5.1. Limitations

The relatively high correlation between pre-experimental aesthetics and pre-usefulness measure was not found for pre-experimental aesthetics and post-usefulness perceptions. System layouts that were considered as aesthetic before use were not perceived as more useful after using the system. A limitation of the current study is that aesthetic perceptions were not measured after using IsraSearch. In future studies, it would be reasonable to examine the correlations of post-experimental aesthetics with post-usefulness and post-satisfaction because aesthetic perceptions may change during and after the actual interaction with a system.

The idea that emotions change the way the human mind solves problems and aesthetics can change our emotional state (Norman, 2004), lead us to expect that aesthetics perceptions will affect usefulness perceptions, user satisfaction and performance - but we did not find this effect. An explanation for the lack of aesthetic effect is that we measured aesthetics by a single construct (see Table 3) .This measurement may not be enough to test the effect of aesthetics by a general aesthetics construct. Aesthetic perceptions are more complex and there may be certain aesthetic dimensions that are more influential than others on other system perceptions and even on performance. For example, alignment and grouping are important for rapid performance (Parush et al., 1998) but not all attractive features of graphic design improve performance (Shneiderman, 2004).

5.2. Conclusions

Norman's idea that emotions change the way the human mind solves problems and aesthetics can change our emotional state (Norman, 2004), lead to the expectation that aesthetics perceptions will affect usefulness perceptions, user satisfaction and performance - but the results show no such effect. Perhaps the aesthetic design did not have a significant effect on the user's emotional state, or it did not affect their emotional state to a point where it affects performance. Norman claims that emotions last for a relatively short periods – even minutes. It will not be unreasonable to think that users in the high aesthetic group felt good for receiving a beautiful interface, and users in the low aesthetic group felt bad for receiving an ugly interface. However, these emotions where very short and relatively weak in the context of a laboratory experiment, allowing them to quickly move their focus to the experimental demands, leaving their feelings behind.

We conclude that it is importance to drill down when investigating the notion of perceived aesthetics. Different system layouts may arouse different aesthetic dimensions. For example, a background of electric wires or chips would give a modern, professional or sophisticated look, while leaves and butterflies would give a very different feel of pleasure, harmony and beauty. Some aesthetic dimensions may influence some perceptions of the system while others may not. For example, Lavie and Tractinsky (2004) found that classical aesthetic dimensions are more closely related to perceived usability than expressive aesthetic dimensions. If the system's "skin" is judged first and creates a halo effect, then system designers should design to arouse the "right" aesthetic impressions. In other words, the

image of the system reflected by its design should fit its purpose. A search engine that is perceived as beautiful, artistic and skillfully designed can at the same time involve elements that are considered as old fashion, and therefore may be perceived as relatively low on usefulness.

We followed previous studies that researched aesthetics of interactive systems, which manipulated aesthetics in terms of changing colors of elements, background texture, font's style, and the location of captions. Subjects were able to express their aesthetic taste and state that one screen is beautiful while another is ugly, but the reasons for these statements may be most important. However, it is necessary to understand why users state that a layout is beautiful and find out whether certain dimensions of aesthetics influence other system perceptions such as usefulness. The results of this research and the results of the studies conducted by Lavie and Tractinsky (2004), show that there is a need to have a better comprehension of aesthetic perceptions of interactive systems. When Lavie and Tractinsky (2004) delved deeper towards a better understanding of the various aesthetic dimensions, their results shed new light on the already known usability-aesthetic relation: perceived usability was correlated substantially higher with the classic aesthetic dimension than with the expressive aesthetic dimension. Perhaps a deeper understanding of the various aesthetic dimensions may similarly reveal whether some dimensions have greater influence on the associations between aesthetics and usefulness. Finer grain perspectives of perceived aesthetics can also follow Hermeren's (1988) distinction between five types of aesthetic qualities: emotion, behavior, gestalt, taste and reaction.

In addition, perhaps different aesthetic dimensions are more influential on the user's general perception of a system, depending on the various contexts of use (such as user' goals and tasks, application genres, etc.). Future research that will view perceived aesthetics in a finer grain resolution and that will take the context of use into account may find that specific aesthetic dimensions have an impact on the perception of a system's usefulness. An interesting research we suggest may test which aesthetic dimensions are more influential on different cognitive processes. Perhaps some aesthetic dimensions are more influential when user's cognitive processes are characterized by automatic behavior (e.g. "freely" browsing the internet with no specific goal), and others are more important when cognitive processes are characterized by controlled behavior (e.g. searching for specific information to accomplish a task with a well-defined goal). Another research route to examine is the possibility that different dimensions dominate aesthetic perceptions of different types of end users (children versus adults, etc.). There are many fascinating paths to follow in understanding the notion of aesthetic perceptions of interactive systems, and its influence on user's attitudes and behavior with those systems.

References

1. Bing, J. and Harvold, T. **Legal Decision and Information Systems**, Oslo: Publications of the Norwegian Research Center for Computers and Law, 1977
2. Chu, H., and Rosenthal, M. **Search engines for the World Wide Web: A comparative study and evaluation methodology**, in Proceedings of the ASIS 1996 Annual Conference, October, 33, pp. 127-35, http://www.asis.org/annual-96/ElectronicProceedings/chu.html (August 19, 2003)

3. Coiera, E.W. and Vickland, V. **Is relevance relevant? User relevance ratings may not predict the impact of internet search on decision outcomes**, Journal of the American Medical Informatics Association 15, (April 2008), pp. 542-545
4. Davis, F.D. **Perceived usefulness, perceived ease of use, and user acceptance of information technology**, MIS Quarterly, 13, 3, 1989, pp. 319-339
5. Davis, F.D. **User acceptance of information technology: System characteristics, user perceptions and behavioral impacts**, International Journal on Man-Machine Studies, 38, 1993, pp. 475-487
6. Floris, van N.-L. **Aesthetics versus utility in electronic imaging**, in Proceedings of SPIE, the International Society for Optical Engineering
7. Hassenzahl, M. **The interplay of beauty, goodness and usability in interactive products**, Human-Computer Interaction, 19, 4, 2004, pp. 319-349
8. Hermeren, G. **The variety of aesthetic qualities**, in: Mitias, M.H. (ed.), „Aesthetic Quality and Aesthetic Experience", Amsterdam: Rodopi, 1988, pp. 11-23
9. Kowalski, G**. Information system evaluation**, in „Information Retrieval Systems: Theory and Implementation", 1st ed. Norwell, MA: Kluwer Academic Publishers, 1997
10. Lancaster, F.W. and Fayen, E.G. **Information Retrieval On-Line**, Los-Angeles, CA: Melville Publishing Co., 1973, Chapter 6
11. Lavie, T. and Tractinsky, N. **Assessing dimensions of perceived visual aesthetics of web sites**, Human-Computer Studies, 60, 2004, pp. 269-298
12. Nielsen, J. **Usability Engineering**, AP Professional, 1993
13. Norman, D.A. **The invisible computer: Why good products can fail, the personal computer is so complex, and information appliances are the solution**, Cambridge, MA: MIT Press, 1998
14. Norman, D.A. **Emotion & design: Attractive things work better**, Interactions of ACM, 9, 4, 2002, pp. 36-42
15. Norman, D.A. **Emotional design: Why we love (or hate) everyday things**, New-York: Basic Books, 2004
16. Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G. and Granka, L. **In Google we trust: Users' decisions on rank, position, and relevance**, Journal of Computer-Mediated Communication, 12, 3, 2007, pp. 801-823
17. Parush, A., Nadir, R., and Shtub, A. **Evaluating the layout of graphical user interface screens: Validation of the numeric computerized model**, International Journal of Human-Computer Interaction, 10, 4, 1998, pp. 343-360
18. Shapira. B, Taieb-Maimon, M. and Nemeth, Y. **Subjective and objective evaluation of interactive and automatic query expansion**, Online Information Review, 29, 4, 2005, pp. 374-390
19. Shneiderman, B. **Designing for fun: How can we design user interfaces to be more fun?**, Interactions, 11, 5, 2004, pp. 48-50
20. Skog, T., Ljungblad, S. and Holmquist, L. **Between aesthetics and utility: Designing ambient information visualiza-tions**, IEEE Symposium on Information Visualization, INFOVIS-2003, pp. 233-240
21. Tractinsky, N. **Aesthetics and apparent usability: Empirically assessing cultural and methodological issues**, CHI-97 Conference Proceedings, Atlanta: 22-27 March, 1997, ACM, New-York, pp. 115-122
22. Tractinsky, N., Katz, A.S. and Ikar, D. **What is beautiful is usable**, Interacting with Computers, 13, 2, 2000, pp. 127-145
23. Tractinsky, N. **Towards the study of aesthetics in information technology**, 25th Annual International Conference on Information Systems, Washington, DC, December 12-15, 2004, pp. 771-780

24. Venkatesh, V. and Brown, S.A. **A longitudinal investigation of personal computers in homes: Adoption determinants and emerging challenges**, MIS Quarterly, 25, 1, 2001, pp. 71-102

25. Venkatesh, V. and Davis, F.D. **A theoretical extension of the technology acceptance model: Four longitudinal field studies**, Management Science, 46, 2, 2000, pp. 186-204

[1] **Acknowledgement**

The author extends her sincere thanks to Liron Rotem for her contributions in designing IsraSearch, planning of the experimental tasks, and running the pilot and the laboratory experiment.