

**This is an electronic reprint of the original article.  
This reprint *may differ* from the original in pagination and typographic detail.**

**Author(s):** Silvennoinen, Johanna; Jokinen, Jussi

**Title:** Aesthetic Appeal and Visual Usability in Four Icon Design Eras

**Year:** 2016

**Version:**

**Please cite the original version:**

Silvennoinen, J., & Jokinen, J. (2016). Aesthetic Appeal and Visual Usability in Four Icon Design Eras. In J. Kaye, & A. Druin (Eds.), CHI '16 : Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 4390-4400). ACM.  
<https://doi.org/10.1145/2858036.2858462>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

# Aesthetic Appeal and Visual Usability in Four Icon Design Eras

**Johanna M. Silvennoinen**

Department of Computer Science and  
Information Systems  
P.O. Box 35, FI-40014  
University of Jyväskylä, Finland.  
Johanna.silvennoinen@jyu.fi

**Jussi P. P. Jokinen**

Department of Computer Science and  
Information Systems  
P.O. Box 35, FI-40014 University of  
Jyväskylä, Finland.  
jussi.p.p.jokinen@jyu.fi

## ABSTRACT

Technological artefacts express time periods in their visual design. Due time, visual culture changes and thus affects the design of pictorial representations in technological products, such as icons in user interfaces. Previous research of temporal aspects in human-computer interaction has been focusing on particular interaction situations, but not on the effects of design eras on user experience. The influence of icon design styles of different eras on aesthetic and usability experiences was studied with the method of primed product comparisons. Affective preferences and their processing times were analysed in order to examine visual usability in terms of semantic distance and aesthetic appeal of icons from different design eras. Aesthetic and usability preferences of icons from different eras varied, which allowed the investigation of the process in which users experience icons. This examination results in elaborating the process, for example the relationship between cognitive processing fluency, familiarity, and beauty.

## Author Keywords

Icon; design eras; aesthetic appeal; visual usability; semantic distance; affective reaction times.

## ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI)

## INTRODUCTION

Due to the development of information and communication technologies, possibilities of various visual design styles for pictorial representations have increased. Advancements in

computer graphics have made varieties of new designs possible, thus increasing the importance of aesthetic design of pictorial representations [43]. Moreover, related to this technological development, the focus of human-computer interaction (HCI) research has shifted from functionality and usability more towards aesthetics of interaction [9, 16, 17, 20] emotional design, [7, 18, 45, 62, 66] and semantic meanings conveyed by pictorial representations [10, 29, 30].

Several studies have reported a positive relationship between perceived usability and visual aesthetic appeal, in which aesthetic appeal increases how well interactive artefacts are cognised [17, 32, 33, 63, 65]. In addition, evaluations of aesthetic appeal are at least partly conducted quickly and these immediate judgments function as critical determinants of first impressions of the encountered artefacts [34, 46, 64, cf. 26, 61]. Additionally, differing dynamics in aesthetic-usability effect have been observed. For example, visual appeal can positively affect the interaction between users and technologies [40], enhance performance in difficult tasks [45], lead to overall decrease or increase in performance, and motivate users to continue using the artefact [58].

The variance and complexity of results related to aesthetic appeal imply that it is a multi-dimensional concept, influenced by factors not yet completely satisfactorily operationalised in HCI research [48]. One disregarded consideration is how different design eras, associated with digital pictorial representations, impact aesthetic appeal and visual usability of user interfaces. Thus, while HCI researchers have studied phases of user experience and, for instance, life-cycles of technological artefacts [e.g. 31, 28, 59], time-related aspects of visual design have not been considered, nor has research focused on the influence of products from different design eras on aesthetic and usability experiences.

Aesthetic valuation is associated with time periods, and each time period defines and refines its value in a relation to previous periods, prevalent design styles and trends [49]. Thus, visual representations help people to reflect on their era. Visual design objects are expressions of a specific era

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

CHI '16, May 07-12, 2016, San Jose, CA, USA

© 2016 ACM. ISBN 978-1-4503-3362-7/16/05...\$15.00

DOI: <http://dx.doi.org/10.1145/2858036.2858462>.

and the existential experience of the people of that era. Further, visual design as an expression of its era helps people to reflect on their time and understand their existential foundations [19]. Constantly changing visual culture thus affects appraisals of pictorial representations. Therefore, the design period of a technological artefact influences how the user experiences it.

## COGNISING THE ICON

### Apperception, aesthetic appeal, and semantic distance

How technology-experience occurs in people's minds can be approached with how the object of our experience is sensibly represented. In this mental process, called *apperception*, new sensory information, as well as already existing mental information, are integrated into a subjectively meaningful and coherent mental representation. Apperception can thus be described as '*seeing something as something*' [22, 27, 50, 56]. For instance, a circle with a line attached to its southwest side (a magnifying glass icon) is associated with 'search' as long as the user has the right mental framework for this interpretation. How this connection is made affects how the user experiences the icon and the overall user interface. If the user is not familiar with the metaphor, she fails to cognise that the signifier (the icon) refers to an affordance ('search') [44]. Further, if the pictorial representation is not perceived as appealing, the user is not satisfied with the design even though she is able to cognise the connection.

This study focuses on how a design era of an icon influences how users experience its visual usability and aesthetic appeal. Aesthetic appeal is operationalised with concepts from traditional aesthetics, and visual usability with the concept of semantic distance. Traditional accounts of aesthetics conceptualise aesthetic responses from the perspective of positive cognitive engagement [1, 11]. Even though, for instance, sadness can be considered as an aesthetic response to an artefact that moves the person emotionally [1]. Positive engagement with pictorial representations includes evaluations of intrinsic value and goodness of the encountered artefact, in which beauty is considered as a good semantic aesthetic property to be experienced [1, 11].

Usability in aesthetic-usability research has been conceptualised with different definitions, such as apparent usability [32], perceived usability [17, 58], and expected usability [60]. The methodological approaches and definitions of perceived usability vary, and the usability evaluations have been conducted by visual inspection of a whole user interface, especially regarding its interactive elements, such as menus and search tools [60]. As usability involves a variety of approaches and definitions, we use the term 'visual usability', by focusing on how people represent visual elements and how the visual aspects of objects relate to the functions afforded by them. Visual usability in HCI refers to how intuitively and fluently visual information can be interpreted in terms of users' goals, and how visual

variables of an interface, such as colours, lines, symmetry and visual rhythm, guide the interaction [54, 55]. In this study, the focus is on visual usability of an icon's semantic distance. In icon design, semantic distance is an important factor in cognitive effectiveness, enhancing icon's processing fluency. *Semantic distance* is the proximity of the icon's pictorial representation to its intended meaning [23, 24, 36, 37, 38, 39 42]. The closer a pictorial representation of an icon can be apperceived to its intended meaning, the easier it is to construct the correct mental representation of the icon's functionality. Icons aim to make information easy and quickly accessible [13, 14, 21, 41]. In this design context, ambiguity is not desirable as the goal is to design efficient visual representations with determined meanings.

### Cognitive processing fluency, appraisal, and familiarity

In addition to the variance of research results concerning the aesthetic-usability effect, ease of interpretation has been observed to predict aesthetic appeal [47, 48]. This has been explained with improvements in cognitive processing fluency by familiarity [48], priming [47], and presentation repetition [1, 67]. Familiarity with pictorial representations eases the interpretation of the encountered stimuli, which has been observed to result in higher ratings on beauty [1, 39, 47, 67]. Priming and repetition cause spreading activation in relevant nodes of the representational network, which affects recall and familiarity [35]. Mere-exposure effect indicates that repeated exposure to a stimulus strengthens the user's affective response to it [67]. However, effects such as mere-exposure may give too narrow insight into the apperceptive process. For example, a congruent design era might affect appeal ratings of familiar stimuli more than familiarity alone, and thus more familiar icons are not necessarily more preferred.

The apperceptive process, in which users evaluate the appeal of products, has been studied as an emotional appraisal process [25, 26]. *Appraisal* refers to the cognitive process, which determines the subjective significance of an event [12, 53]. Because the emotion process does not only involve utilitarian emotions, such as happiness, anger, or fear, but also aesthetic emotions [52], appraisal theory is useful for inspecting how the human mind evaluates objects as appealing [26]. As the process is cognitive, researchers can investigate the processing times associated with aesthetic evaluation, and thus try to elaborate the causes of appeal.

Any cognitive process operates on information, and this is true for the emotional process as well [25, 26, 52, 53]. Emotional appraisal has three information sources: perceptual stimuli, associative processing, and reasoning [57]. In HCI, this notion has explored in terms of evaluation latencies: affective judgments of products take more time as their cognitive demands increase due to the need to invoke different information sources [26]. The associative information source depends on spreading activation, which

means that retrieving relevant information from long-term memory is dependent on the activation of relevant nodes. These depend on repetition, familiarity, and priming, which cause activation spreading [1, 26].

The apperceptive process is rich, and therefore an unfamiliar pictorial representation might not be appraised as appealing, or be apprehended, if an evaluative framework does not exist. Via the cognitive interpretation of emotion, appraisal theory is able to explain why familiarity, priming, and repetition influence aesthetic appeal. In light of appraisal theory, it is possible to explore further the connection between processing fluency, familiarity, and appeal. Any experience of appeal necessarily results from a cognitive information processing. Objects, that are familiar, have lower activation thresholds for relevant long-term memory nodes, and reasoning associated with them has been automatised further (see [3, 51] for accounts of how familiarity relates to long-term memory activation and increased efficiency of reasoning). Thus, one expects more familiar objects to be evaluated faster. While this does not necessarily mean that familiar objects are also more appealing, familiar objects are easier to process, which is desirable in itself. Especially in the context of HCI, where the goals of the user are often related to efficiency, familiarity can thus be expected to correlate with appeal.

Further, because affective judgments are based on a cognitive process that depends on, for example, spreading activation, one should assume that objects that are related to each other, for example by having a common background, should be evaluated within a similar appraisal framework. Thus, for example, artefacts from common design eras should be more similarly evaluated than artefacts from different eras. One can assume that advancements in computing have improved icon design, and thus icons of newer eras are more appealing and have shorter semantic distance, that is, are easier to mentally connect with the affordance they signify. However, the first two hypotheses of this study do not posit the direction for these effects, because users do not necessarily prefer the newest designs before they have gotten used to them.

**H1.** Icon's aesthetic appeal differs between design eras.

**H2.** Icon's semantic distance differs between design eras.

Semantic distance is operationalised as preference and processing fluency. Thus, these two should be correlated. Further, we hypothesise that icons that are closer to their intended meaning, that is, icons that have shorter semantic distance, are preferred over icons that signify the same affordance but have longer semantic distance.

**H3.** Processing fluency correlates with preference.

Design eras are chronological, and do not precede each other arbitrarily. Thus, icons that are temporally distant from each other differ in aesthetic and usability evaluations. Hence, as the temporal distance between two icons increase, their

comparison should be easier and thus occur faster. The appraisal-based explanation would be that icons that are temporally distant from each other do not suppress each other during spreading activation, because they are distant and only vaguely associated.

**H4.** Icons that are distant in time from each other are easier to compare.

The four hypotheses serve two purposes. First, support for them gives plausibility to the use of appraisal theory in understanding how users experience technological artefacts. Such a theoretical basis is welcome in a situation, where user experience research in HCI is inconsistent in its use of such central terms as emotion and aesthetics [4]. Second, psychological exposition of how people experience icons of different design eras gives insight into for example the aesthetic-usability effect, which helps researchers and designers to improve their thinking. Using the results, a designer can consider different design styles and their effect on the desired user experience.

Both purposes presented above serve a further theoretical goal, which is to link together the listed concepts of visual appeal and emotion. The key point is that an aesthetic judgement about a visual object is always partly a cognitive process. While this process is partly non-conscious, it is possible to analyse it using methods from cognitive psychology [26]. In short, the apperceptive process of evaluating visual appeal of pictorial representations is cognitive, which means that it connects information from different sources, processes them, and results in observable phenomena, such as conscious judgments of appeal. In an experiment, one may ask participants to make aesthetic judgements, and by manipulating what is evaluated and recording the processing times, results concerning the cognitive process of aesthetic evaluation can be presented.

## **METHOD**

The hypotheses of this study, as well as the conclusion presented above, imply a controlled study, where observations about behaviour are used to make inferences about the cognitive processes behind that behaviour. In cognitive psychology, this strategy is often implemented in terms of reaction time (RT) studies [2]. The study was thus conducted using the method of primed product comparisons [26]. The idea is to produce both preferences and preference-related reaction times (RT) by presenting primes and stimulus pairs to the participants. Preferences are used to test H1 and H2 related to evaluation of appeal and visual usability, whereas H3 and H4 are tested with RTs.

### **Participants and Stimuli**

The participants were required to be born in 1985 so that each could have at least possibly interacted with the stimuli used in the experiment. Using a mailing list for people interested in participating user psychological experiments, participants ( $N = 30$ ) were recruited. The mean age of the

participants was 35.4 years ( $SD = 5.5$ , range of birth 1964–1985). Half of the participants were women, and half men.

Three adjectives were chosen to measure aesthetic appeal, *beautiful*, *old-fashioned*, and *familiar*. The first two were selected, because from the traditional perspective to aesthetics, experiencing beauty and pleasantness are integral in appraising pictorial representations [e.g. 1, 11]. Also, in line with the previous research, aesthetic is often operationalised in HCI studies by measuring evaluations of beauty [e.g. 16, 65]. In addition, the chosen adjectives are required to be able to measure assessments of visual designs concerning the time of the creation, due to the fact that conceptions of aesthetic appeal alter in time [e.g., 49]. Thus, the prime, 'old-fashioned' allowed us to investigate how the users appraised the effect of time on the icons, which is important when studying design eras. Familiarity refers to the experience that the object is congruent with the subject, and is thus more complex feeling than just recalling how familiar an artefact objectively is. In addition to the three adjectives, three everyday computing functionalities were selected: save, search, and print to study ease of interpretation with semantic distance [37].

In total, 12, icons served as stimuli for the experiment (Table 1). Icons were selected from five different design













	Save	Search	Print
Windows 95			
Windows XP			
Windows 7			
Windows 8			

Table 1. The stimulus icons.

styles: Windows 95 (1995), Windows XP (2001), Windows 7 (2009), and Windows 8 (2015). From each operating system, three icons were selected to match the three functionalities, which were to be used as primes (save, search, and print). All participants had at least some experience with Windows 95, one did not have any experience with Windows XP, three had never used Windows 7, and four reported not using Windows 8.

### Procedure

In the method of primed product comparisons [26], the participants are first shown a *prime* on a computer screen for certain amount of time (here, 3000 ms). Then, the prime is replaced with a *stimulus pair*, and the task of the participant is to indicate, using either of the two RT buttons, which of the stimuli more corresponds to the prime. This choice is called a *preferential match*, which emphasises the fact that any apperceived correspondence is subjective: given an adjective, such as beautiful, the participants prefer one of the two icons. Basically, it is not plausible to claim that appraising something as beautiful would be a result of objectively perceived correspondence, because judgments

of appeal result from an apperceptive process. The participants are asked to make a preferential match as quickly as they can. The participants are also asked to keep their index fingers on the RT buttons at all times during the experiment, except for the four breaks that were distributed evenly throughout the experiment.

One experimental task consisted of a prime, a stimulus pair, and a preferential match, which ended the task and began a new one. All six primes (three adjectives for aesthetic appeal and three functions for semantic distance) were combined with all possible pairs of stimuli, which meant that each participants produced 396 preferential matches with associated RTs (the number of possible pairs from 12 stimuli is 66; this times the 6 primes is 396). The order of the tasks was randomised for each participant.

### Data Analysis

The method of primed product comparisons results in two kinds of data: preferential matches and associated RTs. In order to generalize the preferences of all participants, a *preferential match score* (PS) is calculated. Given a prime and a stimulus, a PS indicates how often the stimulus was preferred over other stimuli with the given prime. PSs close to 1.00 indicate that given the prime, the stimulus was almost universally preferred over other stimuli, and a PS of 0.00 means that the stimulus was never preferred with this prime. For example, the PS of Windows XP search on 'beautiful' was .71, but for Windows 95 save .22, indicating that the former was clearly regarded as beautiful and the latter not. For each of the six primes, 12 PSs were calculated, one for each stimulus.

Further, for each pairwise comparison, a *preference agreement* was calculated, which refers to the fraction of the participants who gave their preferential match to a certain stimulus on a given prime. Agreement scores close to 1.00 indicate that the participants were almost unanimous in their choice of the pair, given an adjective. For example, in comparing Windows 95 search icon with Windows 7 search icon under 'beautiful', only four of the participants preferred the former, giving this pair a high agreement of .87 for 'beautiful'. Conversely, the participants were divided on the beauty-comparison of Windows XP print and Windows 7 print icons (13 versus 17 preferential matches), resulting in a small agreement of .13.

The first two hypotheses were tested by comparing the PSs of icons from different eras. First, mean PSs for each primes were printed for each icon between all eras and visually inspected. To test H1 and H2, One Sample Chi Square tests were conducted for each six primes, with null hypotheses that the four eras were chosen with equal frequencies. Tests for H1 were conducted for the whole data, while tests for H2 considered only tasks in which both stimuli corresponded to the function that was in the prime (e.g., if the prime was 'search', both stimuli had to be search icons for the task to be included in the test). This is because while it makes sense to compare the aesthetic

appeal of two icons even if they do not refer to the same affordance, it does not make sense to do this for functionality. The same thinking was applied in all other calculations that concerned semantic distance.

The third measurement, along with PS and agreement, was RT, which indicates how quickly the participant was able to make their preferential judgment. Mean RTs were calculated for icon pairs, given a prime, and for each individual icon, given a prime. A Pearson correlation between mean RTs and PSs was conducted for each six primes in order to test H3. Large effect sizes ( $r > .50$ ) were taken to support this hypothesis [8]. Correlations between RT and agreement were also calculated for more detailed analysis of H3.

In order to test H4, a generalized linear mixed model was used to predict RT with the age difference of eras as the fixed factor. Age difference was calculated as the difference of the icons on an ordinal scale. For example, the age difference of an icon from Windows 95 and an icon from Windows XP was 1, whereas the age difference of icons from Windows 95 and Windows 8 was 3. The age difference of an icon pair from the same era was 0. H4 states, that larger age difference predicts smaller RTs. In addition to the fixed age difference factor, the model

included a random intercept to account for repeated measures (one individual produced multiple observations of RTs). Further, a Q-Q plot inspection revealed that the RT distribution was not normal, but followed well a gamma distribution with a log-link.

## RESULTS

The mean PSs for the four design eras, between the six different primes are presented in Figure 1. A visual inspection reveals that the effect of design era is not linear for all functions, print save, and search. Whereas icons of Windows XP are clearly more preferred with all three functions when compared to Windows 95, its icons are also more preferred than the icons of Windows 7 and 8. For 'beautiful', Windows 95 is clearly less preferred than the others, which are relatively similarly preferred. Windows 8 is less familiar than the others, and 'Old fashioned' follows a clear age trend. The tests related to H1 and H2 were all statistically significant: 'print',  $\chi^2(3) = 10.1, p = .018$ ; 'save',  $\chi^2(3) = 31.3, p < .001$ ; 'search',  $\chi^2(3) = 17.7, p = .001$ ; 'beautiful',  $\chi^2(3) = 154.5, p < .001$ ; 'familiar',  $\chi^2(3) = 140.0, p < .001$ ; and 'old-fashioned',  $\chi^2(3) = 258.5, p < .001$ . Thus both, H1 and H2 were supported: there were both appeal and semantic distance differences in the icons between the eras.

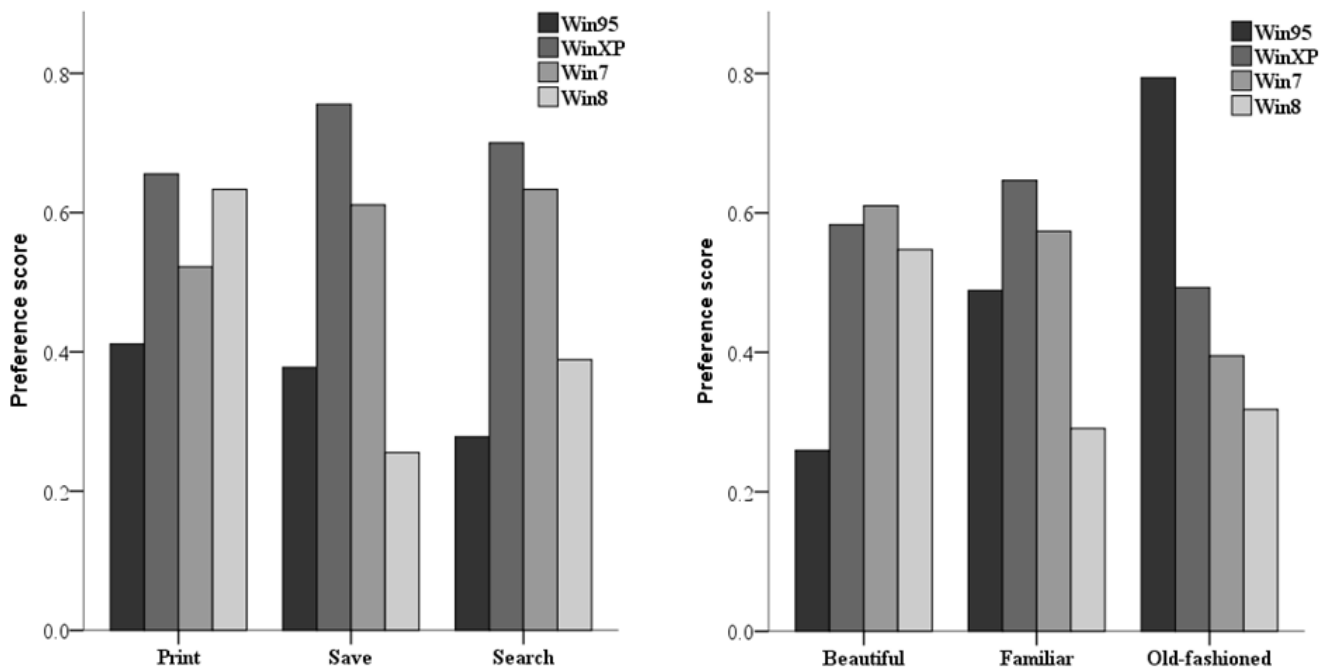


Figure 1. Left: Preference scores for the three functions between the four eras. Right: Mean preference scores for the three affects between the four eras.

	Search	Save	Print	Beautiful	Familiar	Old-fashioned
SD of PS	0.26	0.24	0.25	0.15	0.16	0.20
Agreement	.62	.59	.59	.35	.41	.51
Average RT, ms	1209	1285	1224	1270	1360	1300
$r_{PS,RT}$	-.96	-.93	-.90	-.80	-.47	-.82
$r_{A,RT}$	-.85	-.83	-.86	.11	-.38	-.56

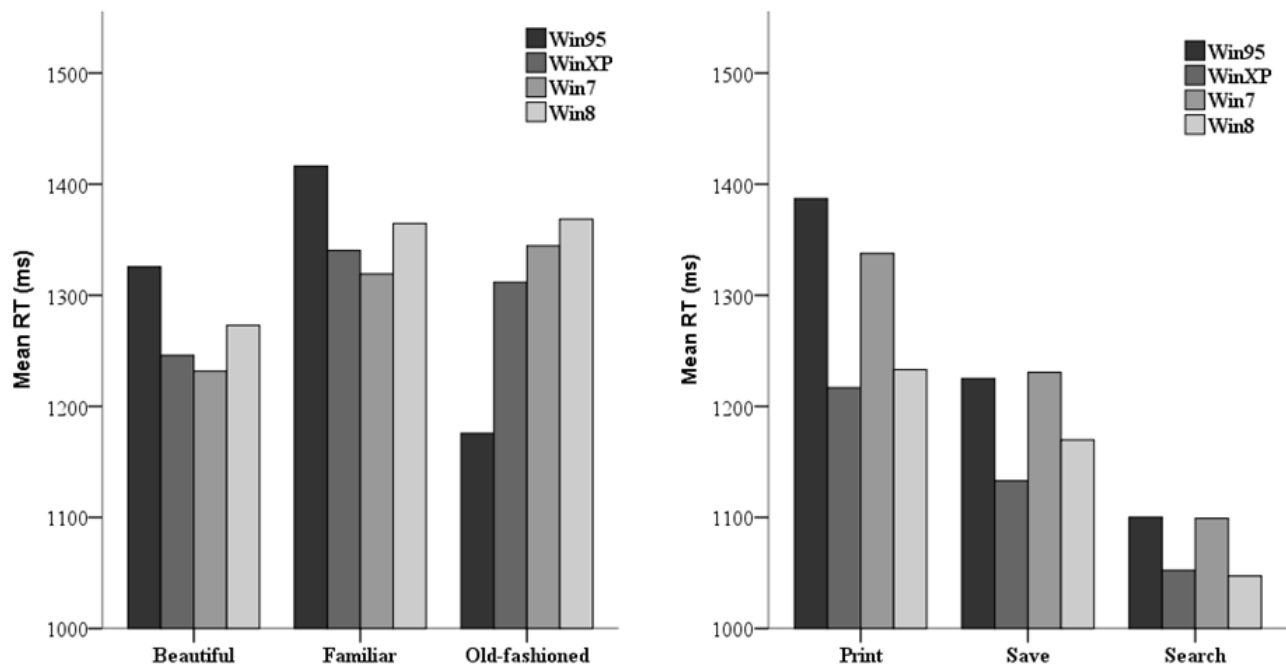
**Table 2. PS spread, agreement, RTs, and relevant correlations from the experiment.**

Mean RTs for the aesthetic appeal affects between the eras are displayed on the left side in Figure 2. Visual inspection reveals that for ‘beautiful’ and ‘familiar’, the RTs were relatively stable, although slightly longer when there was a Windows 95 icon in the comparison. For ‘old-fashioned’, RT was linearly related to the age of the icon: older icons were easier to assess on this prime. Mean RTs for the icon functions between the eras are displayed on the right side in Figure 2. ‘Search’ took, on average, less time than ‘print’ and ‘save’. Further, for all icons, assessments of Windows 95 and Windows 7 icons took more time than for Windows XP and Windows 8, when comparing within the same icon.

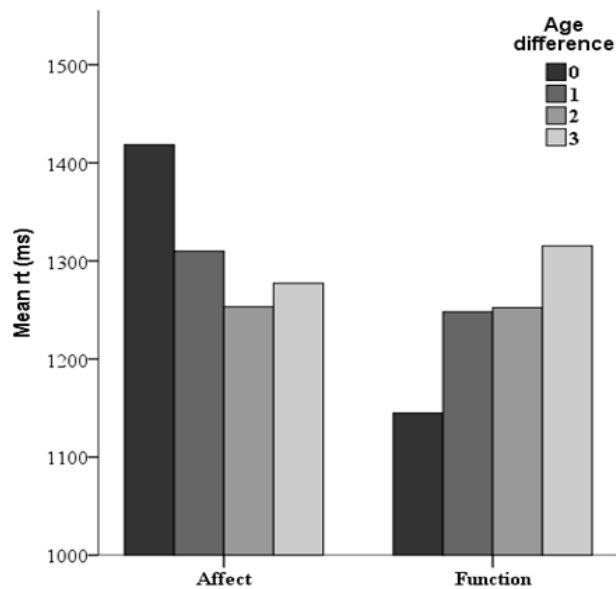
The statistics associated with PS, agreement, and RT are presented in Table 2. Agreement between the participants was greater for semantic distance than for the three aesthetic appeal adjectives: the participants agreed the least

on judgments on ‘beautiful’, whereas there was clear agreement on ‘search’. Similarly, the correlation of RT and agreement is large for the functions, but smaller for the affects. This indicates that for the functions, faster processing time was associated with clearer, more agreed-upon preferences. However, for appeal adjectives, there was less association between processing times and agreement.

The correlations between PSs and RTs ( $r_{PS,RT}$ ) were generally high (.80 or over), giving strong support for H3: processing fluency is associated with appeal. However, for ‘familiar’, the correlation was smaller. The correlations between RT and agreement ( $r_{A,RT}$ ) can be used to investigate H3 further: is ease of interpretation, measured as RT, objective in a way that would result in high correlations between RTs and agreement? The results mostly support this derivative hypothesis, but for ‘familiar’ and especially



**Figure 2. Left: Mean RTs for the three affects between the four eras. Right: Mean RTs for the three functions between the four eras.**



**Figure 3. Mean RTs by age difference, between the prime types.**

‘beautiful’, the correlation does not hold. This is in line with the finding that the participants were not generally in agreement in their beauty judgments.

The multilevel regression for testing H4 revealed, that age difference did not predict RTs,  $F(3, 11876) = 1.8, p = .138$ . Thus, H4 was not supported: icons with more temporal distance were not faster to compare with each other than icons that were temporally close to each other. In order to inspect the result further, a new analysis was conducted, now with a new fixed effect: the type of the prime (appeal or function). Further, the model included the interaction effect between age difference and the type of the prime. In the new model, age difference was still not statistically significant,  $F(3, 11872) = 1.6, p = .199$ , but both the prime type and the interaction effect were,  $F(1, 11872) = 40.6, p < .001$ , and  $F(3, 11872) = 23.3, p < .001$ . Because with age difference of zero, there could be no two stimuli signifying the same affordance, the model was confirmed by including only observations where the stimuli had age difference. The interaction effect was still statistically significant,  $F(2, 9714) = 9.7, p = .012$ .

A closer inspection of the results of the second model for testing H4 reveals that there is actually an effect hypothesised as H4. However, this effect is reverse for the two types of primes. Whereas for functions, larger age difference results in longer RTs, for appeal primes, the direction is the opposite: icons which share the same era take longer to appraise than icons with temporal distance. Mean RTs by age difference, between the prime types are displayed in Figure 3, which shows the interaction effect clearly.

## DISCUSSION

The results supported H1, that there are era differences in icons’ aesthetic appeal. Icons from different design eras were evaluated as beautiful, besides of the icons from Windows 95 operating system. These icons were also considered as old-fashioned and familiar. In icon design research, familiarity has been observed to function as the best predictor of aesthetic appeal, instead of, for example, visual complexity and the level of abstractness of the icon’s visual design [48]. However, experimental examination of the interplay between familiarity and appeal with design era differences indicates that familiarity’s ability to predict appeal ratings is influenced by the design era. Thus, positive valence appraisal is more complicated, for example, advancements in design eras cause Windows 95 icons to be appraised as old-fashioned and non-functional and, thus has negative valence, although considered as familiar. Additionally, icons from Windows 8 were appraised as beautiful but not as familiar. Thus, familiarity and beauty do not necessarily correlate.

Windows 95 icons were consistently appraised as old-fashioned. But since the Windows XP icons, the changes of have been smaller. Although prima facie the changes may be huge, people do not recognise the icons of the eras after the Windows XP that differently. People have a tendency to idolise beauty of pictorial representations of the past [19]. However, perhaps not enough time has passed for appreciation and idolisation of older icon designs. Windows 95 icons were considered as old-fashioned, but not as beautiful. Thus, concerning appraisals of old-fashioned, a design era influence was visible in the favour of newer designs.

The results also supported H2: an era effect was observed in semantic distance for ‘search’, ‘print’, and ‘save’. The icons of Windows 95 and Windows 8 were closest to each other on their semantic distance. This refers to cyclical process of design, in that the design styles of the past re-emerge to current design styles. The similarity between Windows 95 and Windows 8 semantic distance evaluations implies this cyclical design process. From the user’s perspective, icon design has not improved in visual usability since Windows XP icons.

The correlative inspections between RT and PS, and RT and agreement, mostly supported H3, that ease of interpretation is associated with preference. However, some smaller correlations, namely for ‘familiar’ between RT and PS, and ‘beautiful’ and ‘familiar’ between RT and agreement reveal, that this effect is not as straightforward. Familiarity may require more long-term memory and reasoning than the other primes. Thus, although familiarity of a stimulus itself eases mental processes performed on that stimulus, the difficulty of appraising the familiarity of the stimuli is not necessarily dependent on how familiar it is. Thus, appraisal of familiarity is a complex mental process combining memories of encounters and reflection of how congruent the object is with these memories.



Familiarity of an object is not retrieved from memory as much as it is reflected during the encounter, based on what is retrieved and reasoned. This means that familiarity ratings are fluent and context-dependent. This finding can be explained with appraisal theory. A cognitive process *on* a stimulus, such as mental rotation of a three dimensional object, is more fluent with familiar stimuli [5], because familiarity affects the cognitive processes that works on the task. However, a cognitive appraisal process *about* the familiarity of a stimulus still requires relevant information to be retrieved from the long-term memory. This finding suggests that familiarity has different effects on the conscious and unconscious levels of mental processing, and HCI research on familiarity related usability and acceptance effects should consider this possibility.

Although familiarity can often be connected to positive valence, it is not necessarily so, and the appraisal elaboration of experiencing something as familiar is important. The study reported here did not hypothesise association between judgements of familiarity and judgments of beauty, but by correlating the PSs related to these primes, we get  $r = .17$ , or in other words, almost no correlation. Thus, while being able to quickly recognise an object also results in it being considered more beautiful, considering subjectively an object as familiar does not necessarily result in appraising it as beautiful.

Support for H3 justifies the operationalisation of semantic distance as preference and processing fluency. More preferred icons were also faster to appraise, which suggests that these two measures are related to the same phenomenon. This finding suggests that the appraisal process is recursive: first it considers the stimuli from a very general perspective, and if there already are some clear differences, a preferential choice is made. However, if the stimuli cannot be distinguished on this general level, the appraisal process turns to consider the stimuli in more detail. This recursion into increasingly detailed appraisal continues until a judgment can be made, or the task is interrupted. If such an appraisal theoretical perspective into technology experience holds, one should expect more preferred stimuli to be chosen more quickly – exactly what was observed when testing H3. This finding problematises the use of small threshold times in studies of visual appeal in HCI [e.g., 34]. Stimulus exposure times as short as 50ms are not enough for all appraisal process, which leads to the results being not due to the overall appraisal process, but due to just a part of it.

The first test for H4, that longer time differences between stimuli would be observed as shorter RTs, was not supported. However, further inspection revealed an interesting interaction effect associated with H4: while icons farther away in time were easier to compare in aesthetic appeal terms, in functional terms it was more difficult. This suggests that the mental processes associated with affective appraisals consider different aspects than

when evaluating the experience of semantic distance of a signifier to its affordance [44].

People differ in their appraisals of aesthetic appeal, but are more unanimous in evaluating semantic distances between pictorial representations and their intended meanings. A detailed investigation of aesthetics-usability effect gives more in-depth understanding and explanation into the workings of the effect. This requires detailed operationalisation of the concepts of usability and aesthetics with established theoretical foundations. In this study, aesthetic appeal was operationalised with concepts of traditional aesthetics, which enabled detailed analysis. However, aesthetic appeal is a multi-dimensional construct and requires further examination of the underlying constructs, for example, concerning the relationship between familiarity and old-fashioned, and how these function in relation to other dimensions of the aesthetic.

The operationalisation of aesthetic appeal in this experiment provides insight into the aesthetic-usability effect and demonstrates a specific tool for investigating aesthetic appeal and its relationship to usability. Clearly, one dimensional constructs, such as unattractive-attractive [e.g. 34], low-high aesthetics [e.g., 65], non-appealing-appealing [e.g. 58] or pleasant-unpleasant [e.g. 61] are not enough to understand this phenomenon. Appraisal-based conceptualisation of aesthetic appeal used here is not the final operationalisation, but serves as a basis for designing measures and experiments for studying appeal.

The benefit of using the method of primed product comparisons comes from the pairing of stimuli. It reveals that the mental processing associated with product appraisal is more complicated than any straightforward connection between processing fluency and aesthetic appeal. When judging technological artefacts, users require information from different sources, such as perception, long-term memory, and reasoning. These are mental tasks, and thus variance in information requirements results in variance in processing times. Evaluations of aesthetic appeal require more processing time than interpretations of visual usability, due to increased cognitive demands of different information sources.

The chosen method is not meant to replace conventional evaluation, although it does take a specific and important role in studying the mental processes associated with affective judgements. The method allows the study of cognitive processes in more detail when compared to pen and paper questionnaires, as it records RTs. The prime lets the participants make their choices quickly, which allows us to investigate the mental processing times associated with different combinations of primes and stimuli. This is important for theoretical research, as the method permits practical evaluations in the tradition of fast (50ms) evaluations, but, as is shown here, with improved validity.

## CONCLUSION AND FUTURE RESEARCH

The main theoretical and methodological contribution of this study is the analysis of visual usability and appeal as a cognitive process, that is, the clarification of the partly unconscious process, in which people find pictorial representations visually appealing and usable. For example, while icons have overall become more appealing due time (e.g., by comparing Win 95 to Win 7), the detailed effect of eras on icon aesthetics is more complex.

The results of this study provide an in-depth view of the experience process by explicating the influence of design era to aesthetic-usability effect in terms of emotional appraisal, and operationalise the constructs for aesthetic appeal and visual usability in detail. An experiment utilising the method of primed product comparisons revealed that cognitive processing times associated with aesthetic and usability judgments can be used in exploring these phenomena in HCI. The results supported the hypotheses, that the era of an icon affects its aesthetic appeal and its visual usability. Further, the results elaborated the connection between cognitive processing fluency, familiarity, and beauty. Icons that had their intended meaning processed quickly were also regarded as more beautiful, but this relationship did not hold for the experience of familiarity and beauty.

In the study reported here, icons were displayed on a screen without any context, such as other visual elements of a user interface. A better ecological validity would be reached by using complete user interfaces as stimuli. Future studies should thus consider, for example, how the design style affects visual search of icons, when the number of elements on the screen is high. Such research should start to reveal design principles, which can be used to create visually usable and aesthetically pleasing user interfaces in terms of visual design eras.

One interesting consideration from the results of the study is that while icons from the same era create a clearer visual style, the user has to spend more processing to identify relevant icons from a set of stylistically similar icons. In other words, appraising the aesthetic appeal of icons from the same design era requires more processing time, than appraising icons that have time between them. The overall design style needs to be mentally constructed and evaluated, which explicates the processing time due to increased cognitive effort. Evaluating icon's intended functions, larger design era distance requires more processing time, implicating the importance of consistent design style for enabling efficient interaction. In addition, semantic distance evaluations are conducted more unanimously than aesthetic appeal evaluations. Thus, an effective design strategy for improving visual appeal is to focus on semantic distance. Further, as the results show that familiarity doesn't predict aesthetic appeal, but is influenced by the design era of the stimuli, icon design need not necessarily aim for familiarity as long as the era experience is consistent and identifiable. Therefore, the benefit of the approach for design can be restated: conventional design is not necessarily satisfying

design and only right methodology can reveal how we cognise pictorial representations.

The idea of combining processing time and preference in analysing usability and aesthetics of visual user interface elements may prove to be useful in safety-critical contexts. For instance, with in-car navigation user interfaces, being able to quickly separate icons from each other and find the correct one is crucial. Such interfaces are often simplistic, and thus designing icon sets in which the icons are clearly distinct from each other is one of the critical ways to improve the safety via visual design. Reflecting on the results of the study reported here, a designer of such interfaces might wish to consider styles, which mix different design eras. The method of primed product comparison can be used in these kinds of designs to evaluate how distinguishable, but also visually appealing, the icons and icon sets are. In addition, there should be a study in which icons are created or modified according to given recommendations. Other future analyses investigate, for example, the relation of familiarity to the other primes, and the link between agreement and RT for beauty.

## ACKNOWLEDGMENTS

We thank Hilka Grahn and Mikko Jäppinen for helping in the experiment.

## REFERENCES

1. James C. Anderson. 2000. Aesthetic concepts of art. In *Theories of art*, Noel Carroll (ed.), The University of Wisconsin Press, Madison, Wisconsin, 65-92.
2. John R. Anderson. 2000. *Cognitive psychology and its implications* (5th ed.). Worth Publishers, New York.
3. John R. Anderson. 2007. *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press. New York.
4. Javier Bargas-Avila and Kasper Hornbæk. (2011). Old wine in new bottles or novel challenges? A critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '11*.
5. Charles E. Bethell-Fox and Roger N. Shepard. 1988. Mental Rotation : Effects of Stimulus Complexity and Familiarity. *Journal of Experimental Psychology: Human Perception and Performance* 14, 1: 12-23.
6. Robert F. Bornstein. 1989. Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin* 106, 2: 265-289.
7. Susanne Bødker, 2006. When second wave HCI meets third wave challenges. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction* (NordiCHI'06), ACM Press, 1-8.
8. Jacob Cohen. 1992. A power primer. *Psychological Bulletin* 112, 1: 155–159.
9. Antonella De Angeli, Alistair Sutcliffe, and Jan Hartmann. 2006. Interaction, usability and aesthetics:

- What influences users' preferences? In *Proceedings of the 6<sup>th</sup> Conference on Designing Interactive Systems*, ACM Press, 271-280.
10. Pieter Desmet and Paul Hekkert. 2007. Framework of product experience. *Int. Journal of Design* 1, 1: 57-66.
  11. Marcia M. Eaton. 2000. A sustainable definition of art. In *Theories of art*, Noel Carroll (ed.), The University of Wisconsin Press, Madison, Wisconsin, 141-159.
  12. Nico H. Frijda. 1988. The laws of emotion. *American Psychologist* 43, 5: 349-358.
  13. Mariano García, Albert N. Badre, and John T. Stasko. 1994. Development and Validation of Icons Varying in Their Abstractness. *Interacting with Computers* 6: 191-211.
  14. David Gittins. 1986. Icon-based human-computer Interaction. *Int. Journal of Man-Machine Studies* 24: 519-543.
  15. Ravindra S. Goonetilleke, Heloisa M. Shih, Hung K. On, and Julien Fritsch. 2001. The effects of training and representational characteristics in icon design. *Int. Journal of Human-Computer Studies* 55: 741-760.
  16. Marc Hassenzahl. 2004. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction* 19, 4: 319-349.
  17. Marc Hassenzahl and Andrew Monk. 2010. The Inference of Perceived Usability from Beauty. *Human-Computer Interaction* 25, 3: 235-206.
  18. Marc Hassenzahl and Noam Tractinsky. 2006. User experience – a research agenda. *Behaviour & Information Technology* 25, 2: 91-97.
  19. Georg W. F. Hegel. 1835. *Aesthetics, lectures on fine art, vol I*. English translation by Thomas M. Knox (1975). Clarendon Press, Oxford.
  20. Hekkert, P. 2006. Design aesthetics: principles of pleasure in design. *Psychology Science* 48, 2: 157-172.
  21. Shih-Miao Huang, Kong-King Shieh, and Chai-Fen Chi. 2002. Factors affecting the design of computer icons. *Int. Journal of Industrial Ergonomics* 29: 211-218.
  22. Edmund Husserl. 1936. *The crisis of european sciences and transcendental phenomenology*. Northwestern University Press, Evanston, IL.
  23. Sarah J. Isherwood. 2009. Graphics and Semantics: The relationship between what is seen and what is meant in icon design. In *Engineering Psychology and Cognitive Ergonomics, HCII 2009, LNAI 5639*, D. Harris (ed.), Springer-Verlag, Berlin Heidelberg, 197-205.
  24. Sarah J. Isherwood, Siné J. P. McDougall, and Martin B. Curry. 2007. Icon Identification in Context: The Changing Role of Icon Characteristics with User Experience. *Human Factors* 49, 3: 465-476.
  25. Jussi P. P. Jokinen. 2015. Emotional user experience: Traits, events, and states. *Int. Journal of Human Computer Studies* 76: 67-77.
  26. Jussi P. P. Jokinen, Johanna M. Silvennoinen, Piia M. H. Perälä, and Pertti Saariluoma. 2015. Quick affective judgments: Validation of a method for primed product comparisons. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*, ACM Press, 2221-2230.
  27. Immanuel Kant. 1787. *Critique of pure reason*. English translation by Paul Guyer and Allen Wood (1998). Cambridge University Press, Cambridge.
  28. Evangelos Karapanos, John Zimmerman, Jodi Forlizzi, and Jean-Bernard Martens. 2009. User experience over time: an initial framework. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)* ACM Press, 729-738.
  29. Klaus Krippendorff. 1989. On the essential contexts of artifacts or on the proposition that “design is making sense (of things)”. *Design Issues* 5, 2: 9-39.
  30. Klaus Krippendorff. 2006. *The Semantic turn: A new foundation for design*. Taylor & Francis, CRC Press, Boca Raton, London, New York.
  31. Sari Kujala, Marlene Vogel, Anna E. Pohlmeier, and Marianna Obrist. 2013. Lost in time: the meaning of temporal aspects in user experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems- Extended Abstracts (CHI EA'13)*, ACM Press, 559-564.
  32. Masaaki Kurosu and Kaori Kashimura. 1995. Apparent usability vs. inherent usability. *CHI'95 Conference Companion: Mosaic of Creativity*. ACM Press, New York, 292-293.
  33. Gitte Lindgaard and Cathy Dudgeon. 2003. What is this evasive beast we call user satisfaction? *Interacting with Computers* 15, 3: 429-452.
  34. Gitte Lindgaard, Gary Fernandes, Cathy Dudgeon, and Judith Brown. 2006. Attention web designers: You have 50 milliseconds to make a good first impression! *Behavior & Information Technology* 25, 2: 115-126.
  35. Anthony J. Marcel. 1983. Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive psychology* 15, 2: 197-237.
  36. Siné J. P. McDougall and Martin B. Curry. 2006. More than just a picture: icon interpretation in context. In *Proceedings of the International Workshop on Coping with Complexity*, University of Bath.
  37. Siné J. P. McDougall, Martin B. Curry, and Oscar de Bruijn. 1999. Measuring symbol and icon characteristics: norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers* 31, 3: 487-519.
  38. Siné J. P. McDougall, Martin B. Curry, and Oscar de Bruijn. 2001. The effects of visual information on users' mental models: An evaluation of pathfinder analysis as a

- measure of icon usability. *Int. Journal of Cognitive Ergonomics* 5, 2: 153-178.
39. Siné J. P. McDougall and Irene Reppa. 2013. Ease of icon processing can predict icon appeal. In *Human-Computer Interaction, Part I, HCII 2013, LNCS 8004*, M. Kurosu (ed.), Springer-Verlag, Berlin Heidelberg, 575-584.
  40. Morten Moshagen, Jochen Musch, and Anja S. Göritz. 2009. A blessing, not a curse: Experimental evidence for beneficial effects on visual aesthetics on performance. *Ergonomics* 52, 10: 1311-1320.
  41. Kevin Mullet and Darrell Sano. 1995. *Designing visual interfaces: Communication oriented techniques*. Prentice Hall, New Jersey.
  42. Annie W. Y. Ng and Alan H. S. Chan. 2008. Visual and Cognitive Features on Icon Effectiveness. In *Proceedings of IMECS 2008*, 19-21.
  43. David C. L. Ngo and John G. Byrne. 2001. Application of an aesthetic evaluation model to data entry screen. *Computers in Human Behavior* 17: 149-185.
  44. Donald Norman. 1998. *The design of everyday things*. Basic Books, New York.
  45. Donald Norman. 2004. *Emotional design*. Basic Books, New York.
  46. Eleftherios Papachristos and Nikolaos Avouris. 2011. Are First Impressions Only Related to Visual Appeal? In *Proceedings of the INTERACT 2011, Part I, LNCS*, Springer, 489-496.
  47. Rolf Reber, Norbert Schwarz, and Piotr Winkielman. 2004. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review* 8, 4: 364-382.
  48. Irene Reppa and Siné J. P. McDougall. 2014. When the going gets tough the beautiful get going: aesthetic appeal facilitates task performance. *Psychonomic Bulletin & Review* 21, 6. doi:10.3758/s13423-014-0794-z
  49. Richard Shusterman. 2000. *Pragmatist aesthetics: Living beauty, rethinking art* (2nd. ed.). Rowman & Littlefield Publishers Inc., Boston.
  50. Pertti Saariluoma. 2003. Apperception, content-based psychology and design. In *Human Behaviour in Design*, U. Lindeman (ed.) Springer, Berlin, 72-78.
  51. Dario Salvucci and Niels A. Taatgen. 2008. Threaded cognition: an integrated theory of concurrent multitasking. *Psychological review* 115, 1: 101-130.
  52. Klaus R. Scherer. 2005. What are emotions? And how can they be measured? *Social Science Information*, 44, 4: 695-729.
  53. Klaus R. Scherer. 2009. The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion* 23, 7: 1307-1351.
  54. Tania Schlatter and Deborah Levinson. 2013. *Visual usability, principles and practices for designing digital applications*. Morgan Kaufman, MA, USA.
  55. Johanna Silvennoinen, Marlene Vogel, and Sari Kujala. 2014. Experiencing visual usability and aesthetics in two mobile application contexts. *Journal of Usability Studies* 10, 1: 46-62.
  56. Johanna Silvennoinen, Rebekah Rousi, Jussi P. P. Jokinen, and Piia M. H. Perälä. 2015. Apperception as a multisensory process in material experience. In *Proceedings of the Academic Mindtrek Conference 2015*, ACM Press.
  57. Craig A. Smith and Leslie D. Kirby. 2001. Toward delivering on the promise of appraisal theory. Klaus A. Scherer, Angela Schorr, and Tom Johnstone, (Eds.), In *Appraisal Process in Emotion*, Oxford University Press, New York, 121-138.
  58. Andreas Sonderegger and Juergen Sauer. 2010. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics*, 41, 3: 403-410.
  59. Andreas Sonderegger, Gerold Zbinden, Andreas Uebelbacher and Juergen Sauer. 2012. The influence of product aesthetics and usability over the course of time: A longitudinal field experiment. *Ergonomics* 55, 7: 713-730.
  60. Meinald T. Thielsch, Ronja Engel, and Gerrit Hirschfeld. 2015. Expected usability is not a valid indicator of experienced usability. *PeerJ Comput. Sci* 1, e19, DOI 10.7717/peerj-cs.19
  61. Meinald T. Thielsch and Gerrit Hirschfeld. 2012. Spatial frequencies in aesthetic website evaluations – explaining how ultrarapid evaluations are formed. *Ergonomics* 55, 7: 731-742.
  62. Manfred Thüring and Sascha Mahlke. 2007. Usability, aesthetics and emotions in human–technology interaction. *Int. Journal of Psychology* 42, 4: 253-264.
  63. Noam Tractinsky. 2004. Towards the study of aesthetics in technology. In *Proceedings of 25<sup>th</sup> Annual International Conference in Information Systems*, 771-780.
  64. Noam Tractinsky, Avivit Cockhavi, Moti Kirschenbaum, and Tal Sharfi. 2006. Evaluating the consistency of immediate aesthetic perceptions of web pages. *Int. Journal of Human-Computer Studies* 64, 11: 1071-1083.
  65. Noam Tractinsky, Adi S. Katz, and Drori Ikar. 2000. What is beautiful is usable. *Interacting with Computers* 13, 127-145.
  66. Paul van Schaik and Jonathan Ling. 2008. Modelling user experience with web sites: Usability, hedonic value, beauty and goodness. *Interacting with Computers* 20, 3: 419-432. doi:10.1016/j.intcom.2008.03.001
  67. Robert B. Zajonc. 1968. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology* 9, 2: 1-27.