Affordances and Metaphors Revisited: Testing Flat vs. Skeuomorph Design with Digital Natives and Digital Immigrants

David Oswald
Hochschule für Gestaltung
Schwäbisch Gmünd, Germany
david.oswald@hfg-gmuend.de

The arguments about skeuomorph versus flat design have so far been based on comparatively little evidence and were largely dominated by strong opinions voiced in informal online media. This paper presents an a/b-study to assess the strengths and weaknesses of these design approaches with the distinct user groups applying them. First results gave no clear indications of which approach could be considered the best. Surprisingly, however, older and inexperienced users experienced more problems using "naturalistic" skeuomorph interfaces than abstract flat interfaces. As a result, the concept of skeuomorph design enabling the use of real-world knowledge to enhance novice user's learnability is cast into doubt.

Keywords: interaction design, skeuomorphism, flat design, affordance, metaphor, natural user interface, NUI, semantics, digital natives

1. INTRODUCTION

A decade ago we witnessed the revival of strong real-life metaphors in mobile interface design. For five years, so-called "skeuomorph" design dominated mobile interfaces, influencing desktop software and web design. Microsoft's (2011) and Apple's (2013) turn towards "flat design" was followed by an emotional discussion in the HCI and UX community—mainly led on online forums, blogs and non-reviewed online journals. Both sides accused the other of broad ignorance regarding usability questions, and of simply following fashion. [3, 7, 15]

The debate frequently ignores its historic aspect: Firstly, user interface design has a history of "skeuomorph" approaches, peaking in the 1980s with the desktop-metaphor and in the 1990s with multimedia inspired approaches such as Microsoft Bob. Both of these peaks came about when computers were introduced to new user groups. Secondly, we now see a historic shift in user generations. The period when interaction design for adults, all of whom had all grown up in an analogue-only world, was required to be self-explanatory—may soon be over. Today, even two-year-olds use their parents' smartphones or tablets daily (at least in the global north). Hence one is compelled to consider what forms of real world knowledge may now be implemented to facilitate the comprehension of virtual worlds. Clearly such large-scale metaphors as the office/desktop-metaphor would be of no benefit to pre-schoolers. On a smaller scale however, parallels between real and virtual worlds, like "natural" interaction and affordances, may be very useful across all user generations. In this paper, a study design is presented addressing these questions empirically, furnished with the first quantitative and qualitative results.

2. RELATED WORK

Measured by the impact software interfaces have on people's everyday life, academic and empiric studies on visual software design have been relatively rare. However, a comparative study of Windows 7 and Windows 8 [26], demonstrated that the subjects coped far better with the classic Windows 7, containing skeuomorph elements, than with the radically "flat" Windows 8. But, a correlation test with the users' prior experience was not carried out—and the authors admit that Windows 7 did have the advantage of familiarity.

Many contributions explicitly testing skeuomorph against flat design reveal weaknesses in their design. Previous research focused either on small singled out features (e.g. a single icon, a button

style, or different types of gradients) or complex realistic test objects with many simultaneously changing parameters. Furthermore, these tests were largely confined to web pages. User tests of software applications are extremely rare.

Frequently, and most critically, empirical studies in this field are affected by a visual quality far inferior to the industry's state of the art. For instance, in [8] test objects are labelled "flat" and "realistic", where only minor differences are visible. In [16] presenting purportedly "flat" and "skeuomorph" buttons, both examples are made up of black lines and simple grey-scale fillings, resembling early 1990s Windows interfaces. The authors of [9] test an "almost-flat" interface design: The tested objects are abstract circles, filled with a solid grey colour ("flat") and a grey gradient ("semi-flat")—far too simple, if the aim is to apply the results to state of the art software design. This was contrasted by the authors of [6], who tested screenshots of realistic web pages, but the "flat" and "traditional" web pages they presented are different in *every* possible aspect: content, structure, colours, fonts, images, and illustrations. In spite of these confounding variables, the authors claim that the differences they measured are the sole effect of a flat/non-flat design style.

The studies published by the Norman Nielsen Group (NNg) are probably the ones with the biggest impact on the HCI and UX community. However, they have been criticized for a bias against flat design [10]. In [17], the NNg tested two variants of a web page, with identical content and structure, but different UI elements ("flat" vs. "strong 3D-style" buttons). Here too, there was at least one major confounding factor: In the "strong" version, the highlight-colour was reserved for clickable items only, whereas in the "flat" version, the highlight colour was also used for decorative elements. Thus the visual focus was dispersed distractingly over the entire page.

In conclusion here, the goal was to set up a test with a visual design quality close to professional industry standard, and a test design embedding clearly discernable A/B-differences, with a minimal risk for confounding effects, in a realistic context (i.e. use cases).

3. PRELIMINARIES: TERMS AND CONCEPTS

Just exactly what is meant by the terms "skeuomorph" and "flat" design must be explained. Some aspects of these concepts overlap with classic HCI concepts such as metaphor and affordance, whereas other aspects are superficial visual attributes, i.e. style—more akin to marketing than usability issues.

Such confusion may render "skeuomorph" and "flat" ineffective as categories of a scientific discourse. Nonetheless, they are well established terms in an often biased and informal discourse. [3, 7, 15] With this in mind it is worthwhile to sort and classify the diverse aspects of flat and skeuomorph design, and the effects they have on usability. To this end, the following concepts and theoretic frameworks will be employed in the remainder of the paper.

3.1 Affordance

Here, the term "affordance" is used in its original meaning, as introduced by Gibson [12], applied to design by Norman [19], and differentiated by Gaver [11]: the direct perception of possibilities of action and use. Later, Norman complained that the concept has been largely misunderstood and overused. However, his recommendation for HCI to talk only of *perceived* affordances [20], must be scrutinised as affordances have always been defined as a product of *perception* of the environment. Norman's turning away from the concept of affordances to the more general "signifiers" [21] has some disadvantages. Whereas "affordance" describes a specific type of signs (the ones we understand based on causality, i.e. indicators), the term "signifiers" refers to all possible types of signs, thus lacking the required specificity to make detailed sense in HCI. For this purpose semiotics offer many more precise terms and concepts. [22]

3.2 Semiotics and Semantics

Semiotics is a powerful theoretic framework for understanding how meaning is constructed from signs. The semiotic sub-discipline of semantics in particular provides differentiated tools for a deeper understanding of sign usage. Applying the three sign types, index, icon, and symbol are particularly pertinent. [23] The terms differentiate between each other by how the user infers from the sign to its meaning: through causality (index), similarity (icon), or convention (symbol). [13] Defining the sign type by user inference (rather than the visual properties of the sign) implies that the same sign may be understood by seeing a similarity for one user and pure convention by other users: 50-year-olds are likely to understand a telephone icon based on its similarity to the telephone receivers they grew up with. Today's 10-year-olds however, are more likely not to recognise such similarity—as such receivers have vanished from their everyday world. For them, the telephone "icon" is merely a conventional symbol. Furthermore, the concept of index signs (i.e. indicators), which has much in common with the concept of affordance, may help in understanding how meaning is constructed via *perceived causality* in HCI. An introduction to HCI-semiotics by the author can be found in [22].

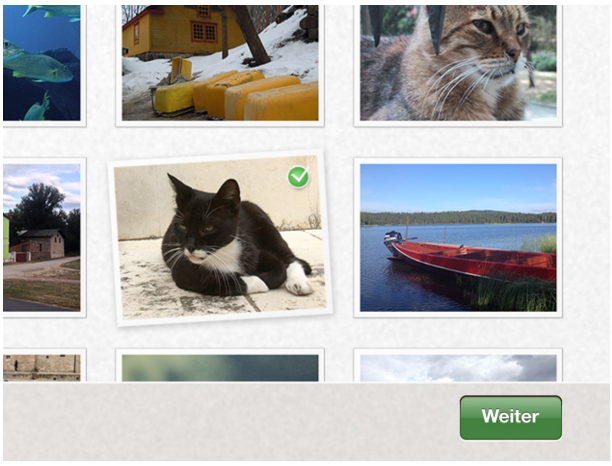

Figure 1. *Skeuomorph "Select Photos" screen (detail), one photo selected. Design by the author.*

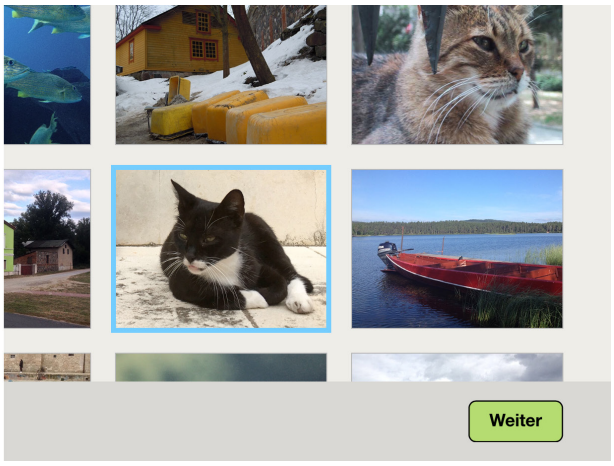

Figure 2. *Flat "Select Photos" screen (detail), one photo selected. Design by the author.*

3.3 Metaphor

The role of metaphors in HCI was hotly discussed throughout the 1980s and 1990s. [4] The purpose for using metaphors was to provide non-expert computer users with interfaces that bear similarity with things they are familiar with. Regarding the first Graphical User Interfaces of the 1980s, the users were office workers familiar with files, folders, desktops, and wastepaper baskets. In the 1990s, other professional software was based on workflows, tools and terminology adopted from analogue predecessors: photographic darkrooms, film flatbed editors, sound mixing consoles, etc. The main difference between metaphor-based interfaces of the 1980s and 1990s to recent "skeuomorph" design is the level of photorealism they display.

According to Apple's 2008 "iPhone Human Interface Guidelines" [1] you should "model your application's objects and actions on objects and actions in the real world. This technique especially helps novice users quickly grasp how your application works." After Apple's turn away from skeuomorphism in 2013, the guiding notion is that "People learn more quickly when an app's virtual objects and actions are metaphors for familiar experiences—whether rooted in the real or digital world." [2] The last sentence is decisive. Recent research [14] reflects this shift from transferring knowledge from the real world to virtual world, to an increasing knowledge transfer from one interface to another—particularly with young users.

3.4 Natural User Interfaces

An interface paradigm also rooted in a real-to virtual-world transfer is "natural interaction". It is based on the assumption, that software is easy to learn and easy to use if it lets us manipulate its objects and its content in a "natural" way, [5, 29] in the best case without any visible handles and controls between user and content e.g. when panning and zooming a digital map on a multi-touch device. However, NUI's biggest advantage—no visible interface—is at the same time its biggest disadvantage. For proficient users, Natural User Interfaces are powerful and fun, but for novice users, without clear visual indicators, they could well simply be confusing puzzles. [25]

3.5 Flat and Skeuomorph Design

In this paper, the term "skeuomorph" is used to describe not only a photorealistic visual style with strong affordances but also an extensive metaphoric approach mimicking real world interaction patterns and workflows. Correspondingly, the term "flat" design is used for interfaces that are not only visually "flat", but also employ indirect ("unnatural") interaction patterns, and an abstract sign language.

4. APPROACH AND METHOD

The study is based on an A/B test design. It is set up as a sequence of small tasks, embedded in realistic workflows. A and B versions differ in their visual design, and partly in their interaction patterns, but they share the same workflow (see tables 1 and 2). A walkthrough video is available online at vimeo.com/270989087 (password: BHCI2018).

Version A employs flat graphics, i.e. solid colours, no shadows, no gradients, no textures nor imitated materiality. Its iconography is preferably abstract and symbolic. Interaction is indirect, using visible controls.

Version B employs skeuomorph graphics, i.e. gradients, shadows, textures, and the like, to create a photorealistic look of physical real life objects. Its iconography is preferably concrete and iconic. Wherever possible, it uses direct interaction with the content, without visible controls.

In order to avoid separated cohorts for A and B tests, two use cases were developed: creating and editing a photo album (for details see table 1), and creating and editing a music playlist (for details see table 2). Hence, each individual participant may run through four test variants:

A: photo album, flat

B: photo album, skeuomorph

C: music playlist, flat

D: music playlist, skeuomorph

Each subject runs through a first test pair, (for instance flat first, skeuomorph second) and a second pair in the opposite sequence (skeuomorph first, flat second). By alternating the sequence of the different scenarios, learning effects from one test variant to the other cancel each other out in the evaluation of task completion time and error rate.

Only four sequences exist with both parameters (flat/skeuomorph and music/photo) alternating after each round: ADBC, BCAD, CBDA, DACB. These sequences were randomly assigned to the test subjects. The tests took place in the participants' private rooms with mobile test hardware and recording equipment. Only the test participant and one observer were present. The test software is a web-based application, run on an Apple iPad 4 in full-screen mode. Click protocols are stored locally and in an online database (screen location and time of clicks, clicked objects and type of interaction). Additionally, the screen and the participants' hands and voice (think-aloud utterances) have been video recorded using a smart phone on a tabletop tripod.

After the test, the participants were asked a set of questions to obtain demographic data (age, gender, educational and professional attainment) and information on their computer and smart-phone use (time and intensity).

The analysis was based on a mix of quantitative (task completion times, error rates) and qualitative methods (error types, trial and error patterns).

Table 1. *Photo Album use case.*

Screen	Instruction text	Task	Flat variant	Skeuomorph variant
1. Home	*Tap on "Photos"*	*Spotting and tapping the icon for the Photos app.*	*Single abstract photo.*	*Pile of realistic photographs.*
2. Select Photos	*Select 5 photos with [a boat / a house] and 5 photos with [fish / a cat]. Then tap "Next".*	*Selecting 10 photos of which only 6 are completely visible, 2 are partly visible, and 2 only after scrolling.*	*Selection indicated by a thick blue frame.*	*Photos with white Polaroid style frame. Selection indicated by off-grid rotation, slight enlargement, hovering effect by drop shadow, and a "checked"-sign.*
3. Edit Photo Album	*Delete the photos on the last page of the photo album. Then tap "Done"*	*Go to the last page of the album. Select the photos. From 3 available buttons, choose Delete (next to Crop and Colour Correction).*	*Abstract album graphics. Page flipping by arrow-buttons. Selection indicated by a thick blue frame. Delete: By X-button.*	*Photorealistic photo album book. Page flip by swiping, with page curl effect. Selection indicated by off-grid rotation, enlargement, hovering effect, and missing photo mounting corners. Delete: By waste paper basket.*

Table 2. *Music Playlist use case.*

Screen	Instruction text	Task	Flat variant	Skeuomorph variant
1. Home	*Tap on "Music".*	*Spotting and tapping the icon for the Music app.*	*Musical notes*	*Photorealistic headphones*
2. Select Music Tracks	*Select 5 tracks with a [blue/x-mas tree] cover and 5 tracks with a [red/snow on the] cover. Then tap "Next".*	*Selecting 10 tracks of which only 6 are completely visible, 2 are partly visible, and 2 only after scrolling.*	*Flat grey track elements. Selection indicated by blue background colour.*	*Embossed, physical track elements. Selection indicated by blue background colour and checked sign.*
3. Edit Playlist	*Move track 5 to the position 2 in the playlist. Delete the last track in the playlist. Then tap "Done".*	*Dragging track 5 to position 2. Scrolling down to the end of the list. From available buttons, choose and tap the last track's Delete-button.*	*Plain lines as drag-element. Delete: X-button.*	*Embossed grooves as drag-element. Delete: Trashcan-button.*

4.1 Sampling of Test Subjects

The number of test participants was limited to 21 subjects. The test subjects were recruited by non-random but quota-corrected convenience sampling. In order to avoid a bias by pure convenience sampling, the aim was to optimize the distributions of gender (10 f, 11 m), age (distribution see table 1), and education (2 elementary school pupils, 2 high school pupils, 1 elementary school, 2 elementary school and discontinued vocational training, 1 high school, 3 middle school and vocational training, 6 bachelor, 2 master, 1 postdoc). To achieve this near even distribution, a convenience sample was complimented with participants recruited via classified ads, offering an allowance equivalent to 12 US Dollars. Thereby, distribution gaps in age, experience, and in education could be filled.

In order to analyse possible correlations between prior experience of users and their performance in the test, a formula to calculate an experience index was developed. It was defined as follows:

$$I_e = ((\sqrt{a_{win}} \times i_{win} + \sqrt{a_{mac}} \times i_{mac} + \sqrt{a_{og}} \times i_{og}) \div a_{gui} + (\sqrt{a_{and}} \times i_{and} + \sqrt{a_{ios}} \times i_{ios} + \sqrt{a_{om}} \times i_{om}) \div a_{mt}) \times 100$$

Experience index (I_e), with user's experience in years with Windows (a_{win}), Macintosh (a_{mac}), other GUI (a_{og}); Android (a_{and}), iOS (a_{ios}), and other multi-touch interfaces (a_{om}); the respective use intensity i, with the values: occasionally = 1, regularly = 2, and professionally = 3. To correct for today's relevance, GUI and multi-touch experience are divided by the time since these UI paradigms have been introduced to a mass market: for GUI a_{gui} = 30 years, for multi-touch a_{mt} = 10 years. The square root is taken to approximate the typical flattening of learning curves over time.

4.2 Analysis

A combination of quantitative and qualitative approaches was used in the analysis. In a first step the software-based click protocols were annotated, i.e. technical information like click location and time stamp were translated into meaningful descriptions (i.e. "tries to drag the song item by clicking on the album cover"). This was done by studying the video footage in parallel. The video footage was also used to check and correct the timing data. For instance, the time needed to spot and tap the Delete-button after scrolling down the playlist had to be corrected manually in case the subjects scrolled down with a "flick"-gesture, which results in the scrolling time being longer than the software-protocolled time between touch-start and touch-end (with the "flick"-gesture, the playlist keeps moving after touch-end, due to momentum-scrolling). The following quantitative measures were calculated:

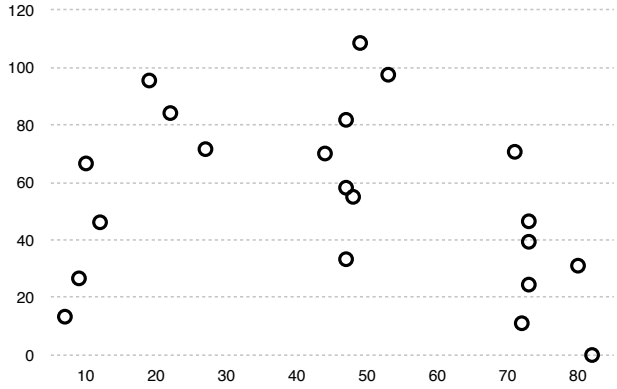

Figure 3. *Distribution of the test participants' age (x-axis) and their experience index (y-axis).*

Home screen: time to spot and tap the "Music" or "Photo" app-icon, prior errors.

Edit Photo screen: time until first successful page flip, prior errors; time between having selected a photo and deleting it, prior errors.

Edit Playlist screen: time until the track is dragged to the other position, prior errors; time between having scrolled down and deleting the last track, prior errors.

General: Total time to complete each of the test variants.

Concerning time and speed, the interest was in the *difference between the test variants* (flat and skeuomorph), as opposed to differences between *individual subjects*. Therefore, first the inter-individual performance-ratio between the flat variant and the skeuomorph variant was calculated, i.e. time of variant A divided by time of variant C of the same subject. In so doing, differences in characteristic personal speed are cancelled out. Only then we compare these individual flat/skeuo-quotients.

In most cases the comparison was done using the median, in some cases additionally by the average. When looking at different portions of the subjects (young vs. old, or inexperienced vs. experienced) also a truncated average was employed, as a compromise between average (too much weight of outliers) and medium (too much deviation between neighbouring values with low number of data points). Statistical tests were done using Spearman's rank correlation coefficient, since a *linear* correlation between the variables is not plausible, i.e. that a subject twice as old (in years) may need more, but not necessarily exactly double the task completion time. Complimenting the timing data, also error rates were counted and compared.

The types of errors were analysed, coded, and compared across all individual subjects in order to build categories and to find reoccurring behavioural patterns. To support this analysis, think-aloud utterances from the video footage were used.

5. RESULTS

The results are organized in six sections and begin with a discussion of aggregated differences of flat and skeuomorph variants (5.1). Subsequently, more specific results will be presented for the photo use case with its differing interaction patterns (5.2) and the music use case with its differing affordances (5.3). Whereas the first three sections offer results that reveal discernable differences between flat and skeuomorph approaches, the results in the remaining sections demonstrate that other aspects may be more influential: First and foremost, the subject's familiarity with other interfaces. This becomes evident in their trial and error strategies (5.4) and in their interpretation of symbols and icons (5.5 and 5.6).

In most cases only a weak correlation was found, for instance between the level of prior experience and task completion times. However, these results still are considered relevant, especially since some of these correlations show a trend to the exact opposite of what would have been expected. For instance, a real-world book metaphor would have been expected to help the understanding of inexperienced users. In contrast, the results show that inexperienced users actually tend to cope better with flat interfaces and their distinct symbols.

5.1 Flat vs. Skeuomorph

Roughly one third of the participants performed faster in skeuomorph test versions (median 23%), the other two thirds in the flat versions (median 25% less task completion time). Looking at subject's age, we see that the oldest third of the participants have more problems with the skeuomorph approach: whereas the youngest and the middle aged thirds make an almost similar number of mistakes in both variants, the older thirds had to try additional 3.75 times before they succeed in the skeuomorph versions (truncated average of the individual differences between the error rate of tests A+C and tests B+D). However, the correlation between age and error rate difference was not significant in statistical tests (Spearman's rank correlation coefficient r = 0.15, with p = 0.54).

When we look at prior experience instead of age, the effect is bigger. The most inexperienced third typically is a mix of the youngest and the oldest subjects of a cohort. This inexperienced third produced 7 additional errors in average (median = 4) in the skeuomorph versions than in the flat versions. Here, statistical tests show a correlation between experience index and error difference of r = 0.43 (Spearman's rank correlation coefficient) with a p-value of p = 0.07.

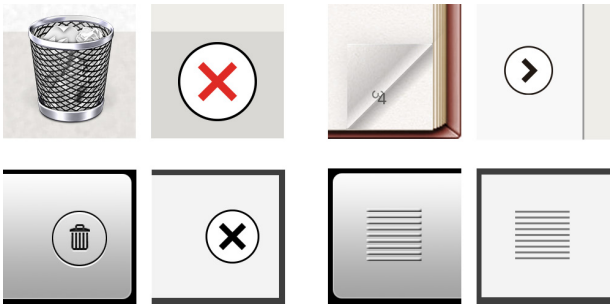

Figure 4. *Skeuomorph vs. flat elements: delete photo (top left), delete track (bottom left), flip album pages (top right), sort music track (bottom right)*

This surprising result suggests that the common idea that skeuomorph real-life metaphors help beginners to understand and learn to use computer interfaces (as Apple [1] suggested for years) is questionable. This will be discussed further in the following section.

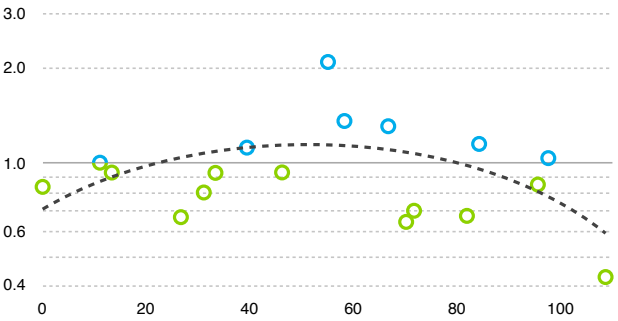

Figure 5. *Ratio of task completion times flat/skeuomorph (A+C)/(B+D) by experience. Blue: subjects faster in skeuomorph version (ratio >1), green: subjects faster in flat version (ratio <1), dotted: trend (polynomial).*

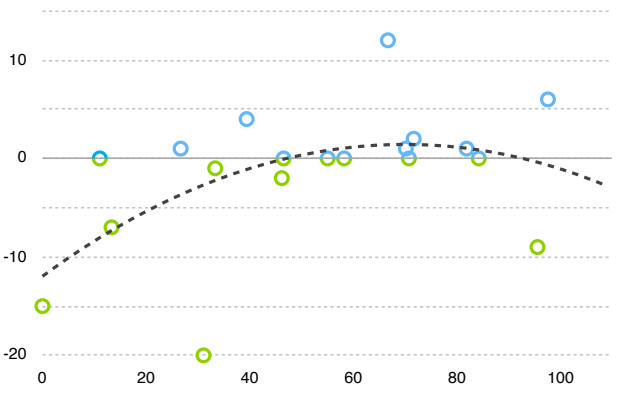

Figure 6. *Difference of errors occurrences between flat and skeuomorph by experience index (1 outlier omitted). Green: subjects with less errors in flat variant. Blue: less errors in skeuomorph variant. Dotted: trend (polynomial).*

Figure 7. *Skeuomorph "Edit Photo Album" screen. Design by the author.*

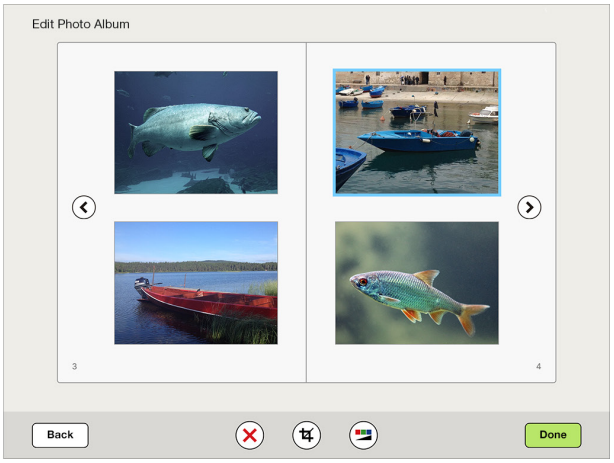

Figure 8. *Flat "Edit Photo Album" screen. Design by the author.*

5.2 "Natural" Page Flip vs. Graphical Next and Previous Buttons

The task with the biggest difference in flat and skeuomorph interaction patterns is "flipping pages" in the Photo Album use case. Whereas the flat variant displays simple arrow buttons for this purpose, the skeuomorph variant indicates that the pages should be turned with a "natural" gesture (a horizontal swipe) by a turned up page corner (see fig. 7). In the test, only the time needed for recognizing the appropriate page turning interaction and starting the first successful page-flip was measured. In this case, possibly different durations of the interaction patterns are irrelevant (tapping a button is in most cases faster than swiping across pages). To avoid any disadvantage for the skeuomorph variant, a simple tap on the (sticking up) page corner as well triggers a page flip.

Figure 9 and 10 show the interpersonal time ratios for first page flips in flat and skeuomorph photo albums. The median time ratio across all 20 subjects is 0.65—which means that the flat arrow buttons were used already after only 65% of the time needed to spot and understand the turned up page corner and start "natural" turning. Or reciprocally: the skeuomorph page flip took 54% more time to understand than the flat version. Only four out of 20 subjects were faster in the skeuomorph version.

An interesting aspect is how the users' performance differs depending on their prior experience with other mobile and desktop software, or with age. Statistically, there is only a very weak correlation for both, clearly not significant, neither with experience ($r = 0.13$, $p = 0.59$), nor with age ($r = 0.24$, $p = 0.31$). Surprisingly however, figures 9 and 10 show trends that rebut the idea that skeuomorph facilitates first use for inexperienced or older users. The trend is exactly opposite.

One reason for the skeuomorph variant's relatively bad results is that inexperienced participants profit much less from so-called "natural" interaction of the skeuomorph book interface than expected. Although it is seemingly obvious that the pages of a book can only be turned in a horizontal direction, especially older participants first tried to scroll *verticality* in the skeuomorph photo album. Two aspects may explain this behaviour: Many inexperienced users have difficulty in understanding that the visual book metaphor also indicates a metaphoric interaction pattern. They do not intuitively infer from the

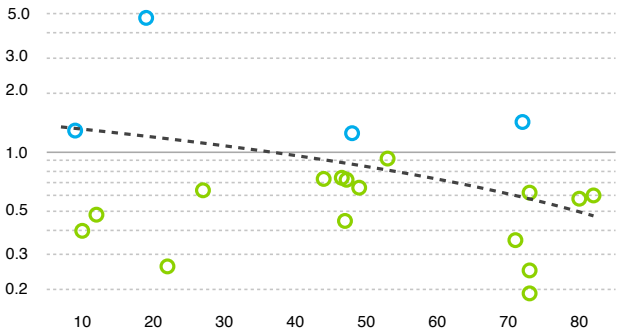

Figure 9. *Flat/skeuomorph-ratio of time until first page flip (y-axis) by age (x-axis), one outlier omitted. Green: subjects faster in flat version. Blue: subjects faster in skeuomorph version. Dotted: trend (linear).*

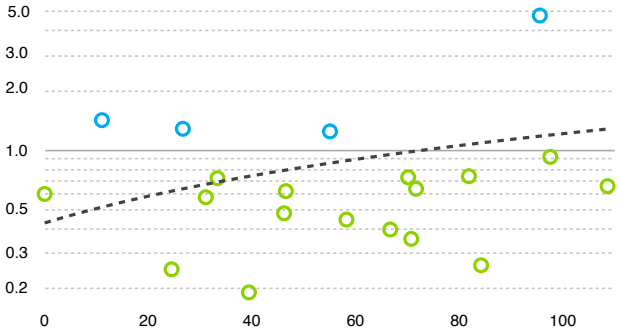

Figure 10. *As fig. 9, but by experience (x-axis).*

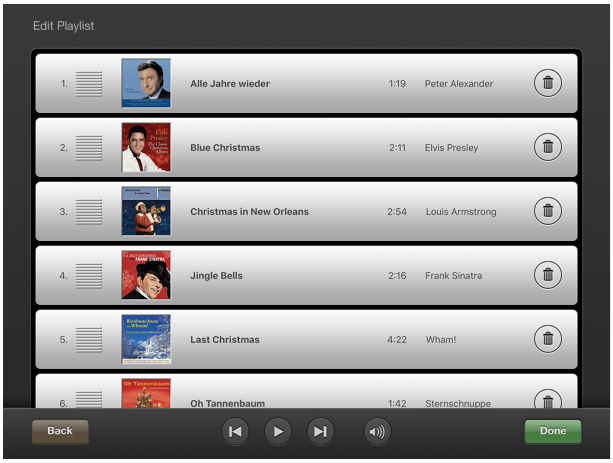

Figure 11. *Skeuomorph "Edit Playlist" screen. Design by the author.*

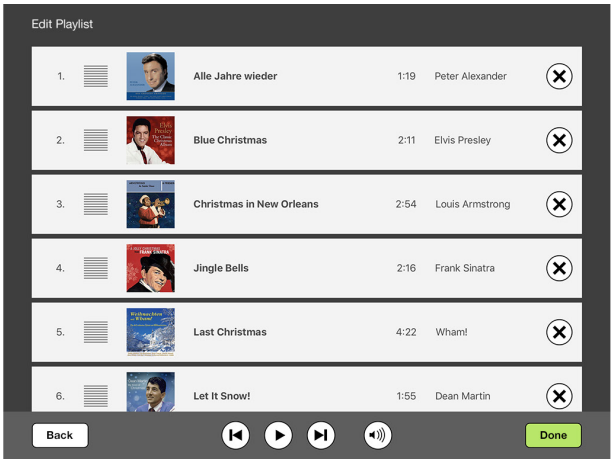

Figure 12. *Flat "Edit Playlist" screen. Design by the author.*

visual real-life book metaphor to the related real-life way of interacting with book pages. On the contrary, the hints by the visual metaphor remain unnoticed, and instead, the first interaction attempts are based on interaction patterns they had been successfully using *before*: On the previous Photo Selection screen (see figures 1 and 2) the scrolling direction was *vertical*. When confronted with the task to go to the last page of the photo album, the first guess on how to do this was based on experience with another interface—even if this experience was the very first and only two minutes old, and although the visual interface did not support the concept of vertical scrolling at all. In consequence, the flat design version of the photo album, with its clearly visible arrow buttons for previous and next pages (see fig. 8) performed clearly better.

A plausible conclusion is that especially inexperienced users profit from clearly visible interface elements that explicitly indicate their functionality. In contrast, experienced users may profit from the ease of use of "natural" interaction principles, without visible handles and controls. Visible controls and buttons are more likely to be self-explanatory and therefore facilitate learning. Direct and "natural" interaction with the content may not be as self explanatory, but in many cases more efficient to use for experienced users.

5.3 Affordances in Flat and Skeuomorph Variants

A concept that shows up regularly in arguments about skeuomorph and flat design is affordance. In regard to skeuomorph design, evidence was found for both, stronger affordances that result in better usability on one hand, but also hints for the danger of misleading false affordances on the other.

In the Music Playlist use case, the equivalent task to page flipping was sorting music tracks in a playlist (rearranging the sequential order). Other than in the Photo Album use case, the flat and skeuomorph variants of the Music Playlist use case are both based on identical interaction patterns. Additionally, both variants make use of the same signifier to indicate the "draggability" of the track items: a grooved area at the left end of the track item, next to the playback number (see figure 11 and 12). This design pattern is borrowed from real-life product design of knobs, switches, and handles that employ grooved surfaces to prevent slipping. Additionally, the orientation of the grooves indicates the direction in which the element may be pushed or turned (direction of force is orthogonal to direction of grooves). This principle has been used in Graphical User Interfaces, for instance to indicate movable scrollbar handles, draggable window corners, and handles for movable panes, etc.

The hypothesis was that a skeuomorph design style would result in a stronger affordance of draggability, due to the more realistic look of the moveable objects. Whereas the grooves are clearly recognizable as such in the skeuomorph version, the plain lines of the flat version may be confused with menu buttons or text symbols – due to the lack of the three-dimensional impression evoked by light and shadow effects. This is confirmed by the completion time measured for the task to "shift track no. 5 to position no. 2", which has been completed 34% faster in the skeuomorph variants (median of all subjects). The weaker affordance of the flat grooved surface also results in a higher number of erratic attempts in the flat variant (median: three times more errors than in skeuomorph). And conversely, the occasions where subjects directly went for the grooved surface in order to shift tracks—without other prior interaction attempts—occurred seven times with skeuomorph and only four times with flat design (in a total of 21 skeuomorph cases + 21 flat cases = 42 cases).

However, the design of the skeuomorph Music Playlist use case also provided misleading false affordances. Several attempts to scroll by swiping over the button bar at the bottom were observed. This occurred twice as often with the embossed skeuomorph button bar than with the flat bar. This observation suggests that skeuomorph design techniques imitating physical affordances should be employed economically and deliberately to indicate existing interaction possibilities only. Apart from providing false affordances, overusing decorative bevel and emboss effects on non-clickable elements may reduce visual focus and eventually will disperse user attention.

5.4 Trial and Error Strategies and Superstition

Insights have also been derived from an analysis of the users trial and error strategies. The task of shifting tracks in the playlist was particularly difficult for subjects who did not associate the grooved area with dragging. Here, it is not only the numbers that are interesting (5.7 trial and error attempts with flat, 2.9 attempts with skeuomorph, in average).

In most cases the first unsuccessful attempts to drag the track element were to tap it somewhere and swipe upwards. However, in so doing, the entire playlist was scrolled rather than the desired single track. Subsequently two different strategies could be observed: 1. Attempting the same interaction, but at a different location of the track element. 2. Trying at the same location, but with a different interaction. In the latter, the most frequent alternative interaction pattern was “longpress”. A very abstract interaction concept one must be familiar with from other interfaces like Android OS.

The “longpress” was frequently attempted in the Edit Photo Album task. In this instance, the subjects had the task of deleting photos. To accomplish this, the photos had to be selected first with a tap, then deleted by a tap on the trashcan icon or an “X”-button respectively. A common unsuccessful attempt to delete was attempting a longpress on the photo, for it to pop up a context menu, which, it was hoped, would feature a delete option. But longpressing the photos, simply made them selected and highlighted—no different to the effect of a simple tap. This led subjects to believe that one *has to* press long in order to select. When selecting the next photo with an unnecessarily long press, this misbelief was confirmed and reinforced. Often this behaviour led to persistent superstition [27], which was actually maintained throughout the rest of the test variants. One subject was convinced that for sorting the playlist items “you have to do it slowly, with patience”, since he was moving slowly and cautiously when he hit the grooved dragging element for the first time. Seeing it worked perfectly, he kept moving very slowly throughout the tests.

Figure 13. *Skeuomorph vs. flat app icons: Photos, Weather, Phone (left), Music, Mail, Clock (right).*

5.5 App Icons

The very first task in each test variant was to spot and tap the respective app icon for Photos and Music. This choice had to be made from six items (see figure 13). Task completion time and error rate were protocoled by the test software. Each of the 21 subjects completed this task four times without any errors in the A, B, C, and D tests. These 84 cases of error-free selection may be interpreted as evidence of an unambiguous and self-explanatory icon design.

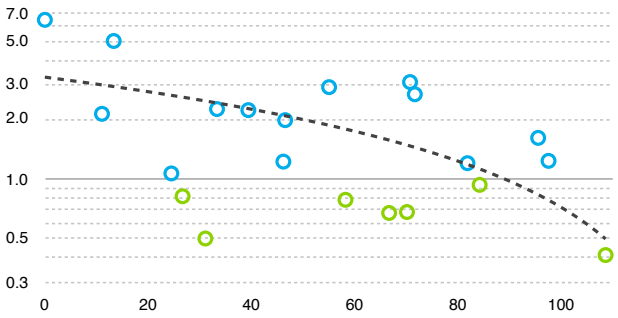

Figure 14. *Flat/skeuomorph-ratio of time to click Photo App icons by experience index (1 outlier omitted). Blue: shorter time for skeuomorph icon. Green: shorter time for flat icon. Dotted: trend (linear).*

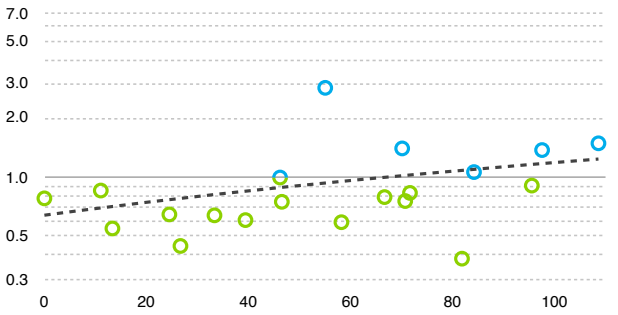

Figure 15. *As fig. 14, but for Music App icon.*

However, the icons showed clear differences in the time it took the participants to recognize and tap them. In the Photo use case the skeuomorph icon performed better (41% less median task completion time), whereas in the Music use case, the flat icon was selected more quickly (21% less median task completion time). This indicates that in this case, the question of "skeuomorph vs. flat" is not necessarily decisive for understandability.

To interpret this result properly, it is necessary to analyse the semantics of all four icon versions in detail. [22, 28] The skeuomorph Photos icon depicts a pile of real photographs, whereas the flat version is reduced to one prototypical and more abstract photograph. In this case, the higher iconicity [18] (more detail and therefore greater similarity between signifier and signified) of the skeuomorph icon leads to better recognition.

The opposite effect can be regarded with the Music icon. Here, the flat version outperforms the skeuomorph version, in spite of higher iconicity, more detail, and a signifier (headphones) that is more closely related to everyday life than the flat musical note symbol. However, the abstract notes symbol outperforms the headphones, because it is generally the more common symbol for music. The Music icon is being decoded primarily based on knowledge from using other interfaces and products, i.e. visual style is clearly outperformed by convention and habit.

Another difference between the Photo and the Music icon is visibility—obviously photographs are visible and signs representing them profit from high iconicity and similarity. Music on the other hand is *audible* and cannot be represented by visual similarity. Therefore, a conventionally known symbol (notes) works better than the depiction of an object, which is only related to music. In simple terms the Photo icon shows photos. The Music icon cannot show music, but symbols or objects *related* to music. This differing semantic "directness" could well explain the differing results.

5.6 Delete: By X or Trashcan?

For the majority of the test participants the interaction sequence for deleting photos was clear. They immediately tapped on the little waste paper basket, or the X-button. This is not necessarily self-evident, since a photorealistic setting with an analogue book and a realistic waste paper basket may also suggest a "natural" (i.e. skeuomorph) interaction pattern: Dragging photos into a waste-paper basket. In fact, this would be much more consistent with the book and waste paper basket metaphor. In the 42 instances of the Photo Album test, only four attempts to drag a photo into the waste paper basket, and one surprising attempt to drag a photo to the X-button, have been recorded. This ambiguity led in average to three more errors dealing with the waste paper basket compared to the X-button (median: 1 additional error). However, it is probably not only the signifier (basket vs. X) that is problematic here, but also the fact that the X is placed on a clearly recognizable button, and the photorealistic waste paper basket is "just standing there". Placing the basket on a button also, may increase the click/tap affordance, and reduce the affordance of "something-can-be-thrown-in-ability". Again, the insight here is that clearly discernable interface elements (like buttons) reduce the ambiguity that skeuomorph interfaces may produce.

6. CONCLUSION AND OUTLOOK

In conclusion it comes as no surprise that there is no simple recommendation favouring either flat or skeuomorph design. The preliminary results of this study already demonstrated that both contain their own pros and cons. Imitating natural interaction promises intuitive understanding. However, as soon as we imitate real world objects and interactions in software design, it remains unclear for an inexperienced user exactly what aspects of the real world concept the designer and programmer have implemented—and which have not. People know that depictions of books on screens are depictions, and that they are different from real books. As a consequence, it is not necessarily clear if swiping or tapping turns the pages—especially for older, inexperienced users. "Unnatural" buttons, by contrast, are more likely to be unambiguous. They are certainly less comfortable for continuous interactions like shifting tracks in playlists—but they are self-explanatory. We find ourselves in a classic design trade-off: Which aspect is more important, unambiguous visibility or direct interaction? Ultimately it depends on the specific use case und the targeted users.

Also when looking at affordances, there is no clear recommendation for flat or skeuomorph design. The example of dragging tracks in a playlist showed that imitating physical objects might make for stronger affordance than flat graphics. However, what classic skeuomorph approaches can build up with one hand may often be torn down by the other: the deliberate strength of one affordance becomes diluted by several other false affordances, that are present only for decorative purposes.

In the next version of the test software, the goal is to sharpen some of the design differences between the variants, and to make the protocol backend more suitable for quantitative testing, thus reducing the burdens of manual data cleaning and manual calculations. With larger sample sizes a more reliable analysis of correlations between learnability and performance, and specific age groups and prior experience should be possible.

7. REFERENCES

1. Apple (2008) iPhone Human Interface Guidelines.
2. Apple (2013) iOS7 Human Interface Guidelines.
3. Bilton, N. (2013) The Flattening of Design. https://bits.blogs.nytimes.com/2013/04/23/the-flattening-of-design (April 9, 2018).
4. Blackwell, A. F. (2006) The reification of metaphor as a design tool, ACM Transactions on Computer-Human Interaction, 13, 4, 490-530.
5. Buxton, B. (1990) The Natural Language of Interaction: A Perspective on Non-Verbal Dialogues. In Brenda Laurel (ed.). The Art of Human-Computer Interface Design, Addison-Wesley, Reading, 405-416.
6. Burmistrov, I., Zlokazova, T., Izmalkova, A., Leonova, A. (2015) Flat Design vs Traditional Design: Comparative Experimental Study. Human-Computer Interaction – INTERACT 2015, Springer, 106-114.
7. Clum, L. (2013) A Look at Flat Design and Why It's Significant. https://uxmag.com/articles/a-look-at-flat-design-and-why-its-significant/ (April 9, 2018).
8. Cho, M., Kwon, S., Na, N., Lee, K. and Suk, H.J. (2015) The Elders Preference for Skeuomorphism as App Icon Style. Proceedings of ACM Conference on Human Factors in Computing Systems CHI EA '15, 899-904.
9. Creager, J.H., Gillan, D.J. (2016) Toward Understanding the Findability and Discover-ability of Shading Gradients in Almost-Flat Design. Proceedings of the Human Factors and Ergonomics Society 2016 Annual Meeting, 339-343.
10. Sean Dexter. 2017. Flat Design: Why you should question Nielsen Norman's research on the trendy design style. https://medium.com/@seandexter1/flat-design-why-you-should-question-nielsen-normans-research-on-the-trendy-design-style-39a991517e02 (April 9, 2018).
11. Gaver, W. (1991) Technology affordances. Proceedings of the SIGCHI conference on Human factors in computing systems, 79-84.
12. Gibson, J.J. (1979) The Ecological Approach to Visual Perception. Houghton Mifflin.
13. Keller, R. (1998) A Theory of Linguistic Signs, Oxford University Press, New York.
14. Jung, H., Wiltse, H., Wiberg, M., Stolterman, E. (2017) Metaphors, materialities, and affordances: Hybrid morphologies in the design of interactive artifacts. Design Studies 53, 24-46.
15. Lomonosov Moscow State University, Laboratory of Work Psychology. 2018. Flat design vs realistic ("skeuomorphic") design: The bibliography. https://www.flatisbad.com (April 9, 2018).
16. Lücken, M., Bruder, G., Steinicke, F. (2015) Evaluation von Buttons im Kontext des Gestaltungsstils Flat Design. Mensch und Computer 2015 Tagungsband, 307-310.
17. Meyer, K. (2017) Flat UI Elements Attract Less Attention and Cause Uncertainty. https://www.nngroup.com/articles/flat-ui-less-attention-cause-uncertainty (April 9, 2018).
18. Morris, C. (1971) Writings on a General Theory of Signs, Mouton, The Hague.
19. Norman, D. (1988) Psychology of Everyday Things. Basic Books.
20. Norman, D. (1999) Affordance, conventions, and design. Interactions 6, 3, 38-43.
21. Norman, D. (2008) Signifiers, not Affordances. ACM Interactions, 15, 6.
22. Oswald, D. (2013) Dynamic Sense-Making in Use Processes of Digital Products – A Semiotic Approach to User Interface Design. Proceedings of AISDR 2013, 5th International Congress of International Association of Societies of Design Research.
23. Peirce, C.S. (1956) Logic as Semiotic: The Theory of Signs, In The Philosophy of Peirce: Selected Writings, Routledge & Kegan Paul, London, 98-119.
24. Prensky, M. (2001) Digital Natives, Digital Immigrants, On the Horizon, 9, 5, MCB Univ. Press.
25. Raskin, J. (1994) Intuitive Equals Familiar. Communications of the ACM. 37, 9 (September 1994), 17.
26. Schneidermeier, T., Hertlein, F., Wolff, C. (2014) Changing Paradigm – Changing Experience? Comparative Usability Evaluation of Windows 7 and Windows 8. Proceedings of Design, User Experience, and Usability DUXU 2014, Springer, 371–382.
27. Skinner, B.F. (1953) Science And Human Behavior.
28. Stickel, C., Pohl, H., Milde, J. (2014) Cutting Edge Design or a Beginner's Mistake? A Semiotic Inspection of iOS7 Icon Design Changes. Proceedings of Design, User Experience, and Usability DUXU 2014, Springer, 358-369.
29. Wigdor, D., Wixon, D. (2014) Brave NUI World: Designing Natural User Interfaces for Touch and Gesture, Elsevier Science, Saint Louis.