

FOR WHAT IT IS WORTH ANTICIPATED EXPERIENCE EVALUATION

Lutz Gegner and Mikael Runonen

Department of Design School of Arts, Design and Architecture Aalto University,
Department of Computer Science and Engineering School of Science Aalto University . I
lutz.gegner@aalto.fi, mikael.runonen@aalto.fi

ABSTRACT

Methods for evaluating concepts on an experiential level are either scarce or very limited in their focus. In this paper we present the Anticipated eXperience Evaluation (AXE) approach for evaluating early product concepts with users. The approach utilizes opposed images as interview stimuli to facilitate the metaphorical thinking and reflection of participants. AXE is a qualitative method that has been designed to provide development teams insights on how future users might experience and value a product or service concept. Furthermore AXE aims at eliciting suggestions from participants that can serve as inspiration for refining concepts. In this paper we provide a detailed description of the AXE approach and its development in three pilot studies.

Keywords: user experience, worth, evaluation, concept, product development

INTRODUCTION

Experiential issues are gaining an increasingly important role in today's consumer product, services or systems¹ development (Pine & Gilmore 1998). Especially in the realm of interactive technologies superior User eXperience (UX) is considered as one of the key competitive factors in product development. We are currently witnessing the expansion of interactive technologies into nearly each aspect of modern life-style. In this highly attractive market, competition is fierce and companies are seeking for innovative applications to increase their market share

and develop new markets. In the area of interactive technologies it is no longer enough for companies to develop products that are highly functional and usable but at the same time they need be desirable for gaining competitive edge. Products are seen much less as tools for completing tasks but more as means to favorable ends (e.g. Cockton 2006)

Human-Computer Interaction researchers and practitioners recognize the limitations of the traditional usability framework to address the non-utilitarian aspects of interaction (Law et al. 2009). Terms like aesthetic (Alben 1996), pleasure (Jordan 2000), emotions (Desmet 2003; Norman 2004), fun (Blythe et al. 2003) etc have entered the primarily pragmatic-oriented community. The introduction of these non-utilitarian aspects asks for new approaches for defining and assessing product qualities.

While the debate in academia is still ongoing what UX means, many usability consultancies have rebranded themselves as UX consultancies and alike while maintaining their traditional methods. Similarly, some consultancies have included their own methods and practices under the umbrella of UX. This careless introduction of the term has led to formation of many different opinions about what UX holds in itself.

It is therefore not surprising that today the definitions and interpretations of UX vary a lot. They range from seeing UX as a mere fashionable replacement of the term "usability" to UX being an inclusive term describing all possible interactions between a person and a company and its offerings akin to "customer experience" (see Law et al. 2009; Gegner et al. 2011). Our perspective of UX follows especially the writings of Hassenzahl and Roto (2007). They distinct UX from traditional usability by considering both: people's be-goals and do-goals. Be-goals are associated with

¹ From this point only referred to as products

needs that are closely connected to people's psychological needs such as being related, competent or autonomous. Do-goals in contrast are instrumental for the achievement of these be-goals and hence focus rather on aspects such as utility, i.e. functionality or ease of use, i.e. the accessibility of functions. This interpretation of user experience shifts the focus from the interaction with a product to the meaning and value of the experience resulting from interaction and product ownership (Cockton 2006; Roto et al. 2009). This view supports the idea that UX design should be an integral part of the entire product development process, hence starting with concept development and not limited to user interface design only.

The focus on experiential aspects during concept development is therefore deemed important to avoid costly corrections in later stages and to increase the probability of success in the market place. One of the major challenges for development teams is the establishment of experiential targets and furthermore the evaluation of concepts when no real interaction in a real context is possible.

In a recent study, Vermeeren et al (2010) collected and analyzed a wide variety of UX evaluation methods. From this collection (available on www.allaboutux.org) the authors identified 24 methods that could be used in the concept phase and 22 on early non-functional prototypes. The methods for concept ideas and early-prototypes are overlapping to a large part, with the exception that within the concept-idea evaluation group inspirational methods for design were included as well. However, almost half of the presented evaluation methods are directed towards experts. Of the remaining a significant proportion are summative approaches and therefore suitable for comparing multiple concepts. Summative evaluation methods are appropriate for identifying the best candidate among a set of concepts or the best combination of features but cannot provide further information or inspiration for refining a concept. Of the remaining methods only a few can be considered useful for assessing the perceived value and meaning of a concept but remain unclear in their description on how the obtained data can be analyzed and interpreted.

CHALLENGES IN EVALUATING CONCEPTS

For the purpose of concept development, design teams should get insightful feedback on how a concept is perceived by a potential target group. Especially experiences triggered by the perceived product character (and its individual features such as functions, content, presentational style and interaction style) (Monö 1997; Janlert & Stolterman 1997; Westerlund 2002; Hassenzahl 2003) provide teams valuable insights for optimizing the UX. This step does not only help in identifying potential problems early in the process but as well extend market and user research by reframing the concept space.

The abstract nature of concepts is one of the main challenges, which makes it difficult to gain valuable feedback from potential users. The presentation of a concept to participants inevitably determines the feedback one can obtain in an evaluation session. Van den Hende (2010) argues that an important characteristic of the presentation method should be that it stimulates the participant's imagination, hence to make him or her envision the future value of the concept in question.

In order to familiarize the participant with the concept, an overall concept description is usually provided to familiarize the participant with the main characteristics of the concept. Also, early concept narratives (van den Hende et al. 2007) are used to translate an abstract concept into a concrete use situation.

Typically an early concept narrative follows a classical storyline with an introduction and an end, describing the use situation of the product as well as its benefits and its attributes.

Simple visual concept representations or early prototypes should be used to illustrate the product concept and to manifest the concepts as an "object" to which participants can refer during an evaluation session.

It is well known that people perform poorly when trying to describe their future needs (Von Hippel 1986). This poor performance is underlined when experiential aspects are taken into account. Describing experiences with words is a difficult task in itself and making a user to imagine a fictitious experience makes it even more challenging. If a person is asked to imagine and explain an experience, there will be a bias caused by the interviewer's wording. Another problem arises with answers elicited from participants

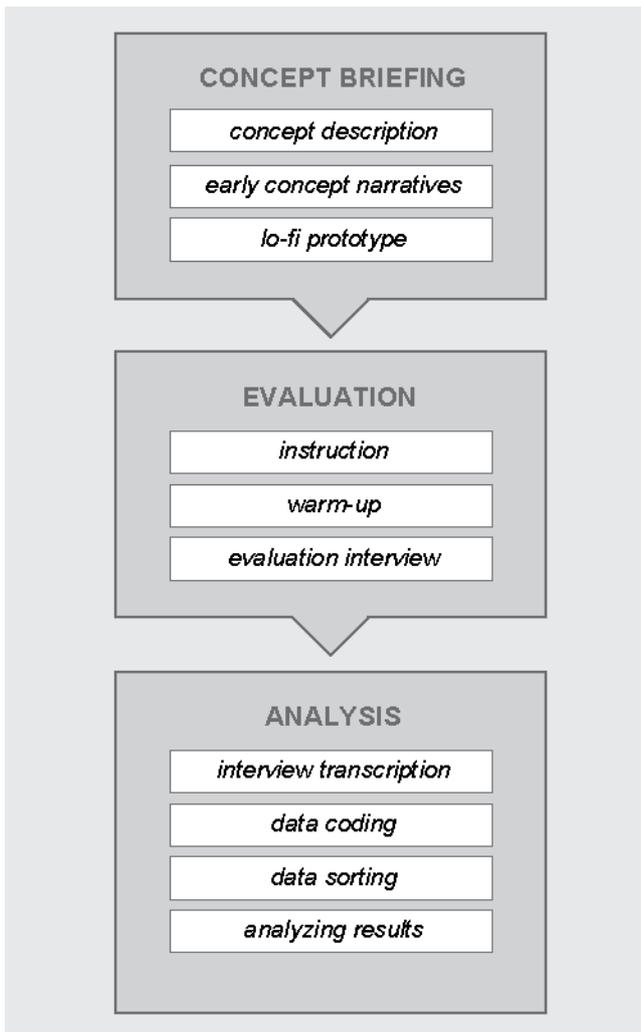


Figure 1: Evaluation process

as words can be interpreted in endless ways. The meaning behind individual words can be derived to some extent from the context. This approach, however, is very vulnerable to misinterpretations and subconscious bias due to the desire of seeing the evaluated product in a positive light by the person doing the analysis. Words in general, both written and uttered, convey poorly subjective appraisal. Especially adjectives need explaining for understanding both, the meaning of the adjective and whether it contains a positive or negative judgment. For example, the sentence “*This is a funny product, it makes me laugh*” can mean a multitude of different things. Firstly, does “*funny*” refer to something that enables a person to have fun or is the product itself ludicrous? Secondly, is “*funny*” in this setting a positive judgment? For a situation that is serious, a funny product may not fit and thus create a negative outcome.

In the following we will introduce our approach that addresses the above-mentioned challenges. The approach supports design teams in gaining insights on the perceived value of concepts by utilizing image-pairs as stimuli in user interviews.

METHODOLOGY

The AXE approach can be divided in three major steps: concept briefing, concept evaluation, and data analysis (see figure 1).

Prior to the evaluation the design team has to define the design targets against which the results ought to be compared to. The establishment of design targets gives the development team a shared understanding of the goals throughout the development stages and the ability to assess whether user’s perception of the concept matches the goals.

CONCEPT BRIEFING

At the beginning of the evaluation session the concept needs to be presented to participants each time in the same manner and order to guarantee comparable results. The description is read out loud by the facilitator to the participant while the participant is also handed a copy so that he or she can return to description at a later stage to clarify the perception.

The early concept narratives are presented to the participant in a similar manner by reading them out loud and providing a copy. In order to make the story more engaging and personal, second-person narratives are used to avoid participants speculating about experiences of fictional characters rather than theirs. If the narratives are accompanied with other material than text (e.g. illustrations, or lo-fi prototypes), these materials are handed to the participant so that he or she can use them freely throughout the session.

CONCEPT EVALUATION

The main element of the AXE method comprises of an answer form consisting of image pairs and scales (see figure 2). The purpose of the form is to give each interview a similar structure and to both help and steer participants to talk about the experiential aspects he or she perceives. The answer form consists of three different parts: *Instructions*, *warm-up* and *image pairs*. The *instructions* are meant to familiarize the participant with the evaluation approach and guide

through the process. The instruction is kept separate from the actual evaluation form, so returning to it is easier for the participant. The instruction page also contains one warm-up exercise to practice the evaluation with the participant. The warm-up gives the facilitator a chance to find out if the participant understood the procedure correctly and spot possible problems. The warm-up pair has been chosen to be highly contrasted and therefore easy to go through. This should encourage the participant so that he/she feels comfortable continuing with the following image pairs.

Image-pairs

The type of challenges mentioned normal interviews and questionnaires entail could be decreased to some extent through generative or enabling methods. Generative methods such as cultural probes (Gaver et al. 1999), make-tools (E. B. Sanders 2001) or context mapping (Visser et al. 2005) employ physical or visual aids such as disposable cameras, diaries, collages, etc. to allow people to express their experiences, expectations and dreams. Enabling techniques have been widely applied in market research and typically employ ambiguous stimuli to allow research participants to express thoughts and feelings towards products that are otherwise hard to articulate (Boddy

2005). In recent years, generative and enabling (sometimes also referred to as projective techniques) have been increasingly emerging in the area of UX research as well (e.g. Kujala & Nurkka 2009; Sproll et al. 2010).

The approach described here utilizes image-pairs as stimuli to aid participants in reflecting and expressing their experiences, attitudes, opinions and beliefs towards a given product concept. The deployed image-pairs are composed to display a contrast and linked through a scale to strengthen the idea of bipolarity. The scale is therefore not a measuring scale but an aid to assist participants during the evaluation interview to better express their perception of the product concept and to indicate their preferences. The image-pairs used in AXE were inspired by AttrakDiff2™ (see www.attrakdiff.de) to have a sound basis for covering the main experiential aspects. The AttrakDiff2™ questionnaire comprises of semantic differentials (e.g. pleasant – unpleasant, good – bad) for evaluating and comparing the pragmatic and hedonic quality of interactive products. Pragmatic qualities correspond to usability and usefulness whereas hedonic qualities are subdivided into dimensions of identification and stimulation. Attractiveness serves as an evaluative construct to

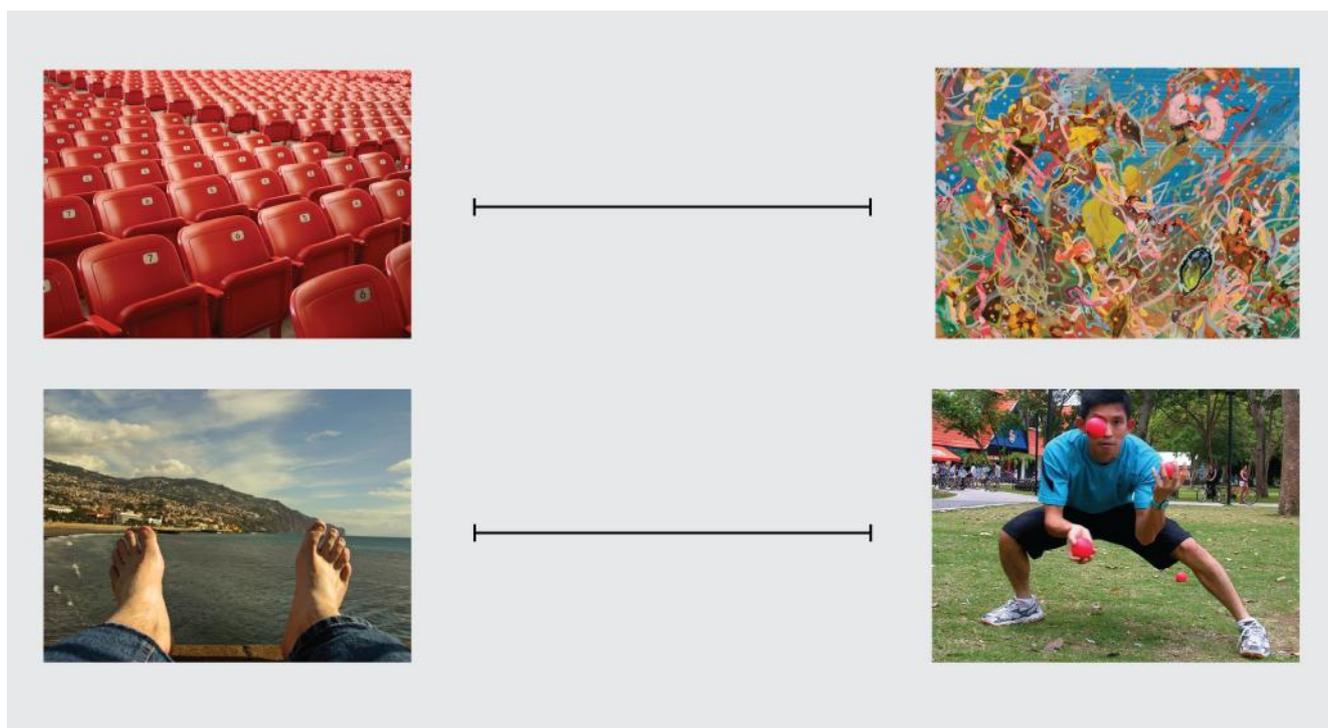


Figure 2: examples of image pairs: The top pair aims at triggering responses towards pragmatic qualities (structured – unstructured) The bottom image pairs aims towards triggering responses on stimulation (undemanding – challenging).

assess the perceived beauty and goodness.

The images were selected to represent the individual anchors visually.

In the following, the process of selecting the image pairs is sketched out in a few steps:

- brainstorming for concrete or abstract depictions that might carry the desired contrasting metaphorical meaning for a particular item (e.g. pleasant – unpleasant)
- retrieving images following a set of exclusive criteria: no professional images (e.g. professional image databases); avoidance of bias concerning gender, age, or race; no cut-out images (i.e. where the background has been removed)
- combination of images with similar qualities (i.e. color/black and white, abstract/concrete, etc...)
- testing image-pairs with a small set of people.

In total the AXE tool comprises of 12 items plus two blank pairs that can be filled by the development team to address experiential issues that are of particular interest for the concept at hand (e.g. trust, security, etc...)

Evaluation interview

During the evaluation session the participants are asked to indicate which of the presented images he/she more closely associates with the concept in question. The made choices serve as a platform for the facilitator to start a conversation around the concept. One natural way of starting a conversation around a pair is to ask why the participant associates the concept more with image A instead of image B. The idea behind this question is to reveal the self-constructed semantic differential the participant is using in his/her decision. As the participant explains his or her choice, the facilitator asks continuation questions that probe deeper into the choice. Furthermore the facilitator asks the participant to indicate on the answer form her/his preference to get a clear understanding of the participant's evaluation of the concept.

If the participant has expressed that the concept is closely associated with A but would rather prefer B, the facilitator can ask why she/he prefers B to gain better understanding of the participants expectations towards the concept. Furthermore this allows the

facilitator to probe which changes in the concept would be appreciated to obtain suggestions from the user on how to improve the concept. To better illustrate the process an exemplary interview is presented in the following:

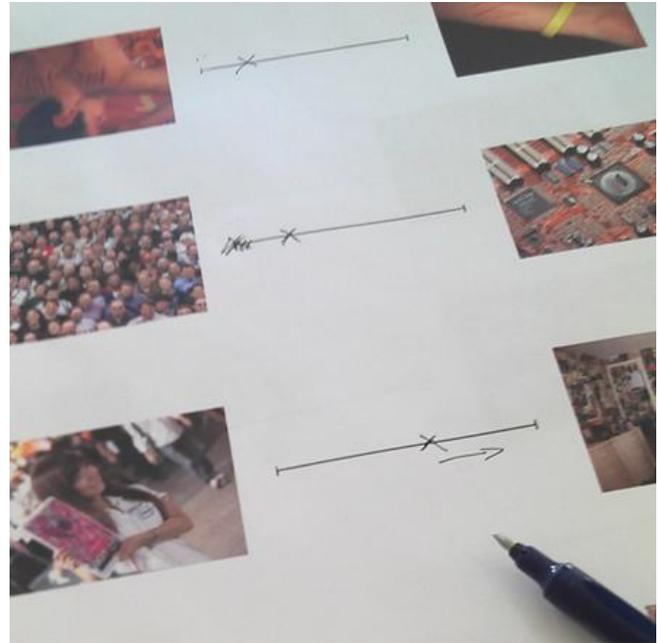


Figure 3: Evaluation form with markings

The participant has indicated that image B is more closely related to the concept. After a first inquiry the participant explains that she/he has evaluated the concept based on "Natural-Unnatural" scale as response to the image pair. The participant has also marked that the concept is strongly towards Unnatural but Natural is preferred. The concept is a drawing application for a tablet pc.

Facilitator (f): *"You have marked, that you associate the concept strongly with 'unnatural'. What makes you feel it is unnatural?"*

Participant (p): *"The way how it works is unnatural to me."*

(f): *"Can you tell me in a more detailed manner what makes 'the way it works' unnatural?"*

(p): *"I guess it's because how you draw with your finger on the screen. I guess it could get more natural in time. At the moment it still feels very unnatural to me compared to using pen and paper how the screen responds to my finger movements."*

At this point the facilitator has extracted from the participant that he or she perceives the interaction as

unnatural. The facilitator could continue and ask for example “*What is so unnatural about the way it responds to your finger movements?*” or approach the issue from a different angle by asking “*What makes using pen and paper more natural for you?*” to obtain a better understanding of “natural” characteristics of drawing for the participant.

In the interview, the facilitator should pay extra attention when asking questions. All the questions should base on the information the participant has provided him/herself. The facilitator must avoid using adjectives that have not been used by the participant, as there is a high risk that the participant assumes the word and a question bias is caused.

DATA ANALYSIS

Subsequent to the interview, the data is transcribed and analyzed. In the following the individual steps in the analysis are described in detail.

Transcription

The analysis of the session data requires a transcription. The transcription has to be done word-to-word in order to preserve as much of the information as possible. The valuable data for an image pair generally starts from the participant choosing the association and ends when moving to the next pair.

Selecting segments

Once the transcription is complete, the text is partitioned into more manageable segments. Each segment, or a snippet, should carry singular observations only. An observation can be, for instance, an expression of attitude towards the concept, claim about a function or a comparison between two attributes. All snippets should be coded according to the source.

Example:

(p) “*This image reminds me of calmness. I don’t think this concept is very calm. If it didn’t have such flashy colors or if there were no animations, it would be calmer.*”

When the example is partitioned into snippets, there are four different observations:

- (1) “*This image reminds me of calmness.*”
- (2) “*I don’t think this concept is very calm.*”

(3) “*If it didn’t have such flashy colors, it would be calmer.*”

(4) “*If there were no animations, it would be calmer.*”

The last sentence carries two different items and therefore it has to be broken down into two different snippets. In order to maintain the connection between the subject and the appraisal, “*it would be calmer*” has to be duplicated. Since the first snippet obviously does not carry information about the concept but only about the image that was used for extracting information, it can be dismissed.

Analytical framework

After a segment has been extracted, it is coded to categorize the information it is carrying.

The categories applied for analyzing the data have been adapted from Hassenzahl’s UX Model (2003) and are illustrated in the analytical framework (see figure 4)

The main categories’ classes reflect the current state of the concept and comprise of perceived product features, associated attributes and anticipated consequences. Additionally a fourth category class carries information for improving the concept within the product context domain. In the following the individual categories are presented in more detail.

Perceived product features are the triggers for associated attributions and consequences within a certain context. In the following the individual sub-categories in this category class are explicated in more detail:

- **General** captures statements, which cannot be attributed to a particular product feature but are directed at the concept as a whole.
Example: “I don’t feel like I have to have it.”
- **Content** refers to the activities that are enabled or supported by the concept.
Example: “I think playing tic-tac-toe is great fun.”
- **Functionality** relates to functionalities the concept comprises.
Example: “I didn’t understand why there was an option to print.”
- **Interaction** refers to issues concerning the operational use.
Example: “I think flicking through the data sheets using swipe gestures is really fast.”

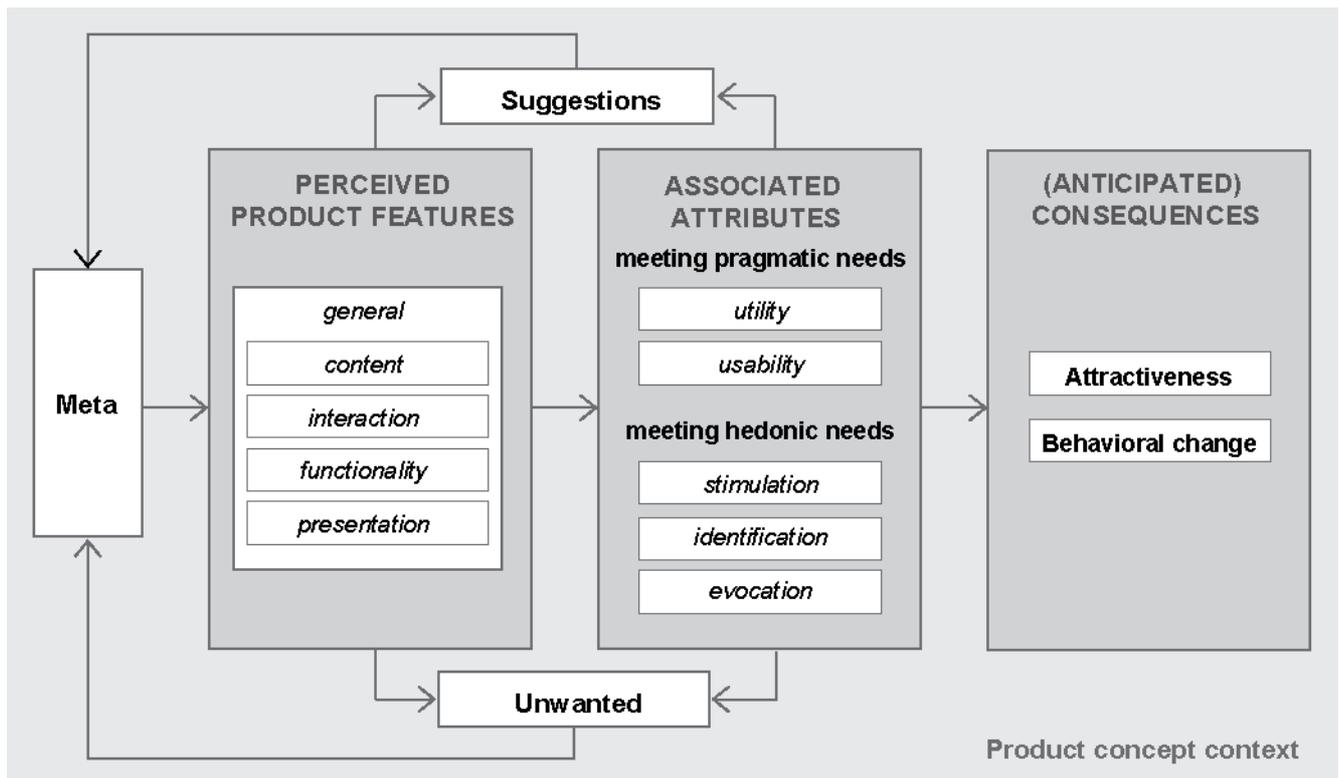


Figure 4: Analytical framework

- **Presentation** refers to the look and feel of the concept
Example: "Those rounded corners were quite ugly."

Associated attributes are summarized in the second category class and refer to impressions participants connect with the product concept.

- **Pragmatic attributes:** Concept's perceived capability to support the user in executing a certain task. This category covers issues like utility and usability.
Example 1: "I think it's pretty much organized."
Example 2: "The PDF-function is very handy".
- **Stimulation:** Concept's perceived capability to motivate use. This dimension covers aspects such as perceived novelty, learning, engagement, excitement and fun.
Example 1: "I would be excited when I use the application."
Example 2: "I haven't seen such a gadget before."
- **Identification:** Concept's perceived ability to communicate a favorable identity to relevant others and how participant can relate to the concept.

Example 1: "If I post a nice picture but the comments are not positive, it might hurt."

- **Evocation:** Evocation refers to the concept's ability to provoke participant's memories. The concept communicates past events, relationships or thoughts that are important to the individual.
Example: "But seeing an emoticon would make me realize of the memories that have passed long back."

Anticipated consequences represents the third category class, which classifies mainly judgments about the concept and comprises of two subcategories:

- **Attractiveness:** This category captures the participant's perceived consequence of the concept or a particular characteristic.
Example 1: "It is a very enjoyable way to express my feeling."
Example 2: "This concept satisfies your desire to get connected with others"
Example 3: "The interface is not so nice."
- **Behavioral change:** A concept may introduce a change in behavior for the participants. Consider for example a speech-to-text input for text messages used in public places. The descriptions

of change and attitudes are to be documented here. It is also possible that the participants describe voluntarily how using a concept or product could change their behavior.

Example: "I'd listen to more music if I had this product."

trashing them can be beneficial in situations where the evaluated concept is part of a bigger totality, for instance a product family, and thus they can have later uses.

Example (in context of hand-held devices): "I don't like to carry much stuff with me when I'm jogging."

The fourth category provides information for potential improvements, i.e. expectations and desires, and enhanced knowledge about current practices.

- **Suggestions** present either conditions such as if "X was present in Y then I would use Y" or straightforward expressed suggestions for improvement. Suggestions represent a particular form of experiential quality, i.e. expectations, desires. However suggestions need to be handled carefully as they cannot be taken at face value for improving the design but need to be carefully interpreted. Also suggestions can be either targeted at the general concept or a particular product feature and can therefore be classified accordingly.

Example: "There should be a back button."

- **Unwanted** refers to perceived adverse outcomes which captures statements about product features or attributes that are not appreciated.

Example: "That confirmation screen annoys me so much"

- **Meta** captures comments that "fall in between", in the sense of not addressing the concept or linked activity directly but still carry interesting information. Collecting these snippets instead of

Ideally, each observation carries a subjective valuation of something being desirable or undesirable, good or bad, positive or negative (attractiveness). In the straight-forward case valuation can be deduced directly from the snippets. If this is impossible, valuation can be done indirectly by utilizing surrounding snippets. The answer sheet can prove to be a valuable asset for marking perceived valuation because the participants mark their preferred options. If the snippet in question is such that a) valuation does not make sense or b) valuation cannot be marked with great certainty, it should be marked as not applicable. An ideal case of coding is shown in *figure 5*.

Following the coding of the data, the individual snippets are sorted based on perceived product feature, associated attribution and by positive, negative and not applicable evaluations. This allows the development teams to quickly identify the perceived strengths and weaknesses of the general concept and its individual features.

Comments within the fourth category class concerning suggestions or unwanted are linked with product features, whereas *Meta* comprises information that is independent of the concept and therefore does not establish links to other categories.

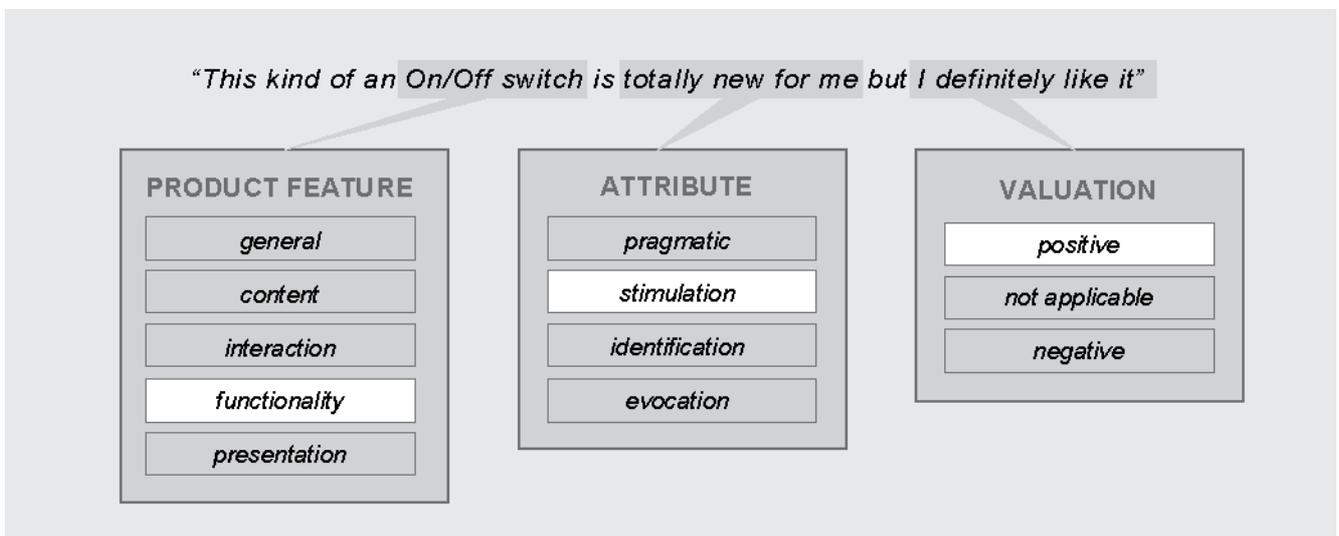


Figure 5: Coding of an ideal example.

DEVELOPMENT AND REFINEMENT OF AXE

Three pilot studies have been conducted to validate and refine the earlier described method. In the following the individual pilot studies are described and the evaluation setups explicated. Furthermore observations and subsequent modifications of the approach will be presented. Please see figure 6 for a summary of the most important items and modifications for each iteration.

PILOT 1

The first pilot study was conducted with two alternatives of an innovative social networking service. Eight students (6 master students, 2 doctoral students) from the department of Computer Science at Aalto University could be recruited (seven male, one female). The participants were mainly of Finnish origin (7). The participants received a movie voucher as compensation.

Evaluation Setup

As we were exploring new grounds the main purpose of this first study was to test whether our participants would respond to the image pairs in the way it was conceived by the authors. Therefore the sample was split into two groups to test two different conditions. The first group was asked to evaluate the concept based on 20 image pairs that had been selected beforehand by the authors. The images had been selected to represent items in AttrakDiff2™ to increase the probability to cover the majority of

relevant experiential qualities. For the second group the 20 image combinations had been completely randomized to increase the ambiguity of the presented stimulus.

Subsequently to presenting the concept the evaluation principle was explained to the participants. Two images had been arranged on a piece of paper with space in between. After presenting the image pair the participants were asked to create adjectives based on the images that would represent a semantic difference. Afterwards we asked them to express which of the adjectives is more closely related to the concept and how strong that association was. Furthermore the participants should indicate which of the two extremes was desirable in relation to the concept from their point of view. The following interview was following the same principle as described in the method section.

Observations

The first study clearly showed that the participants were able to create semantic differentials on the majority of image pairs. Furthermore it quickly became apparent that the participants could express their thoughts about the concept freely and in a very reflective manner.

The two conditions showed that even though it was possible for the participants to make sense even of the randomized image pairs, it took them considerably more time and effort. Furthermore once they understood the evaluation principle they started to deviate from the “script” and followed an approach

	PARTICIPANTS	IMAGE PAIRS	MODIFICATIONS
PILOT 1	8	20	<i>include a scale, pre-selected image-pairs only, questionnaire type of format Added categories: Suggestions, Meta</i>
PILOT 2	8	16	<i>Removal of semantic differential elicitation, preference instead of polarity, Added categories: Behavioral change, Content</i>
PILOT 3	6	12 + 2	<i>Itemless scale instead of 7-point, focus pairs Added categories: Unwanted, Evocation</i>

Figure 6: Summary of modifications between different pilots.

that was more natural to them.

One adverse effect of the adjective elicitation was the participant's urge to explain the semantically opposed adjectives to the facilitator. The interpretation of the selected picture pairs matched only in a few cases directly the intentions of the authors. However, the reflective talk triggered by the image pair, nevertheless addressed in most cases the intended experiential dimensions. Overall the collected data provided rich and deep insights on the way people perceived the concept.

PILOT 2

The results of the first pilot were encouraging to further pursue and refine the approach. Based on the learning of the first study, it was decided to modify the tool in various ways. First of all a 7-point scale was introduced to visually link the images and for strengthening the notion of polarity. Furthermore the results pointed towards using images that had been pre-selected in order to decrease the cognitive workload of the participants and increase the effectiveness of the tool. Results of the first trials as well showed a saturation of the responses after 20 pairs, therefore the total amount of pairs was reduced 20 to 16. As a next step, the image pairs were not introduced individually but as multiple-items in a questionnaire type of format. Analysis for the first pilot also showed that there was a need to add two categories. Participants gave suggestions for improvement very willingly and these suggestions could not be comfortably fitted to existing categories. Also, we identified that some information wasn't connected with the evaluated concept but could have uses later on. *Meta* was introduced as a new category to be used for such information.

Evaluations

In the second pilot the MindMap concept by Lucero et al. (2010) was used for testing a refined version of the AXE approach. Eight students (Four master students, four doctoral students) from various cultural backgrounds (Chinese, Dutch, Finnish, Indian, Italian(2), South Korean(2)) from the Department of Design at Aalto University participated in the second study. All participants received a movie voucher as compensation. The concept was introduced by using a general concept description, two narratives, and a

video demonstration to illustrate the main interaction principles. Evaluations were carried out in an isolated room that allowed participants to relax and go through the process without interruptions.

OBSERVATIONS

Despite the reduction of evaluation items, the overall time needed for an individual session did not decrease. Furthermore, items that had been working well in the previous study caused problems in the second one. The fact that various items were visible at the same time did not lead to irritations or distraction, but rather allowed the participants to determine the speed and the order of the evaluation themselves. The more flexible evaluation procedure also allowed the facilitator to focus more on questions and probe on interesting aspects that were brought up by the participant.

The introduction of the scale proved to be very valuable for strengthening the idea of polarity and evoking reflective talk. An interesting observation was that the identified differentials were in some cases neither perceived positive nor negative, but rather the space between the two extremes was considered desirable. Overall the participants talked freely and it did not require too much effort for the facilitator to keep the discussion ongoing.

Analyzing the results for a concept meant for creating mind maps proved somewhat difficult as some participants were quite indifferent about the concept but expressed strong opinions about the activity of creating mind maps. Also, in some cases, they stated that the nature of the activity would change if such a concept was used.

PILOT 3

The elicitation of the semantic differentials proved to be challenging and time consuming in the course of an evaluation session. In order to streamline the approach and reduce stress for the participants we decided to leave out this step and rather encourage spontaneous ratings. Also, in some cases participants expressed that they don't necessarily see one option (image) more preferable than the other. This led us to rewrite the instructions so that the participants would mark on the scale the preferred value instead of choosing between the two. Furthermore it was decided to reduce the overall image pairs to 12 and

leave room for two additional pairs that would allow the development team to focus on specific topics that should be particularly addressed.

To solve the problem of disconnected activity and product perception, two new categories were included: *Behavioral change* to capture how a concept would alter an activity connected to it and *Content* to differentiate between the concept and the enabled activity.

Evaluations

In the third pilot study a novel photo sharing application was evaluated for testing the modified version of the AXE tool. Six master students (Chinese (1), Finnish (3), Indian (2)) from the Department of Computer Science at Aalto University could be recruited. All participants received a movie voucher as compensation. Evaluations were performed in a usability lab.

Observations

Evocation, a part of the experience framework on which AXE builds on, was deliberately left out formerly as the assumption was that memories could not be addressed sensibly in an evaluation session with a new concept. This assumption was proven wrong in multiple occasions. As photos serve often as memorabilia, many participants formed comments that could not be coded easily with the current framework. These comments were however valuable and clearly connected to overall perception of the concept and could thus not be overlooked. This resulted in re-including *Evocation* as one of the attributes.

While coding the transcripts, it came clear that some opinions were massively more significant than others. This was particularly true with some negative statements. This led to introducing a new category, *Unwanted* as a means to highlight and mediate very undesirable things to developers.

As the semantic differentials were removed, it didn't make much sense anymore to have a quantified scale between them. The scale was replaced with a line to which participants could make their markings.

Because of the nature how the concept is introduced with AXE, there is a possibility that some of the participants misunderstand the functions or the whole purpose. While testing the method, we found out that this problem could be reversed as a possibility. The

error should not be corrected by the facilitator as soon as the misunderstanding is found but explored further if the errors do not compromise the whole evaluation.

The facilitator can try finding out through questions what makes the participant think the function or the purpose is such or such and collect valuable information about how the concept is perceived.

DISCUSSION

Experiential aspects of products are elusive, both because they are difficult for users to express and they are difficult for researchers to interpret. The space of experiences is vast and it can be studied from many different angles. Some of the difficulties can be attributed to the use of language itself and some to the abstract nature of both concepts and experiences. Currently, there is an abundance of methods for studying users' experiences with products but those addressing concept-level products are rare. We have created AXE to provide a method to study how people perceive a product (concept) at a very early stage and to help developers in refining and steering it further.

Using visual stimuli as a starting point for evaluating a concept has proven itself a very viable option. Our participants were able to find meaning to concepts and relate to future use and produce usable information. The analysis framework for AXE was formed through multiple iterations. The base of it is constructed on earlier findings (Janlert & Stolterman 1997; Westerlund 2002; Hassenzahl 2003; Monö 1997) but extended with new articles to better suit concept development needs.

We used testing environments comparable to a lab setting. A logical next step in developing the method further would be taking it to a real context. In case of more radical concepts, the context should be one intended by its developers whereas a more incremental one should use a context existing with similar products.

ACKNOWLEDGEMENTS

We would like to thank Guido Grassel, Virpi Roto, Andrés Lucero, Hannu Korhonen, Mika Rautava and Yanqing Cui for their support and the Nokia Research Center for funding this research.

REFERENCES

- Alben, L., 1996. Quality of experience: defining the criteria for effective interaction design. *interactions*, 3(3), pp.11–15.
- Blythe, M.A. et al., 2003. *Funology: From Usability to Enjoyment* 1st ed., Springer.
- Boddy, C., 2005. Projective techniques in market research: valueless subjectivity or insightful reality. *International Journal of Market Research*, 47(3), pp.239–254.
- Cockton, G., 2006. Designing worth is worth designing. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*. Oslo, Norway: ACM, pp. 165–174.
- Desmet, P.M., 2003. Measuring emotions. *Funology: From Usability to Enjoyment*. Kluwer, Dordrecht, the Netherlands.
- Gaver, B., Dunne, T. & Pacenti, E., 1999. Design: Cultural probes. *interactions*, 6(1), pp.21–29.
- Gegner, L., Runonen, M. & Keinonen, T., 2011. Oscillating Between Extremes: A Framework for Mapping Differing Views on User eXperience. In *Designing Pleasurable Products and Interfaces 2011*. Designing Pleasurable Products and Interfaces 2011. Milan, Italy: ACM.
- Hassenzahl, M., 2003. The thing and I: understanding the relationship between user and product. In M. A. Blythe et al., eds. *Funology: from usability to enjoyment*. Kluwer Academic Publishers, pp. 31–42.
- Hassenzahl, M. & Roto, V., 2007. Being and Doing - A perspective on user experience and its measurement. *Interfaces Magazine*, 72, pp.10–12.
- van den Hende, E., 2010. *Really New Stories The Effect of Early Concept Narratives on Consumer Understanding and Attitudes*. Delft, The Netherlands: Delft University of Technology.
- van den Hende, E.A. et al., 2007. Using early concept narratives to collect valid customer input about breakthrough technologies: The effect of application visualization on transportation. *Technological Forecasting and Social Change*, 74(9), pp.1773–1787.
- Von Hippel, E., 1986. Lead Users: A Source of Novel Product Concepts. *Management Science*, 32(7), pp.791–805.
- Janlert, L.-E. & Stolterman, E., 1997. The character of things. *Design Studies*, 18(3), pp.297–314.
- Jordan, P.W., 2000. *Designing Pleasurable Products* 1st ed., London, UK: Taylor & Francis.
- Kujala, S. & Nurkka, P., 2009. Identifying user values for an activating game for children. In *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*. MindTrek '09. New York, NY, USA: ACM, pp. 98–105.
- Law, E.L.-C. et al., 2009. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the 27th international conference on Human factors in computing systems*. Boston, MA, USA: ACM, pp. 719–728.
- Lucero, A., Keränen, J. & Korhonen, H., 2010. Collaborative use of mobile phones for brainstorming. *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pp.337–340.
- Monö, R., 1997. *Design for Product Understanding: the Aesthetics of Design from a Semiotic Approach*, Stockholm, Sweden: Liber.
- Norman, D., 2004. *Emotional Design: Why We Love (or Hate) Everyday Things*, New York, US: Basic Books.
- Pine, B.J. & Gilmore, J.H., 1998. Welcome to the experience economy. *Harvard Business Review*, 76, pp.97–105.
- Roto, V. et al., 2009. User Experience Evaluation Methods in Product Development (UXEM'09). *Human-Computer Interaction-INTERACT 2009*, pp.981–982.
- Sanders, E.B., 2001. Virtuosos of the Experience Domain. In *Proceedings of the 2001 IDSA Education Conference*. IDSA Education Conference.
- Sproll, S., Peissner, M. & Sturm, C., 2010. From product concept to user experience: exploring UX potentials at early product stages. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. pp. 473–482.
- Vermeeren, A.P.O.S. et al., 2010. User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. NordiCHI '10. Reykjavik, Iceland, pp. 521–530..
- Visser, F.S. et al., 2005. Contextmapping: experiences from practice. *CoDesign*, 1(2), pp.119–149.
- Westerlund, B., 2002. Form is function. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*. DIS '02. New York, NY, USA: ACM, pp. 117–124.