Measuring the dynamics of remembered experience over time

Evangelos Karapanos [a,*], John Zimmerman [b], Jodi Forlizzi [b], Jean-Bernard Martens [a]

[a] Eindhoven University of Technology, Department of Industrial Design, P.O. Box 513, 5600 MB, The Netherlands
[b] Carnegie Mellon University, Human–Computer Interaction Institute and School of Design, Pittsburgh, PA 15213, USA

ARTICLE INFO

ABSTRACT

A wealth of studies in the field of user experience have tried to conceptualize new measures of product quality and inquire into how the overall *goodness* of a product is formed on the basis of product quality perceptions. An interesting question relates to how the perception as well as the relative dominance of different product qualities evolve across different phases in the adoption of a product. However, temporality of experience poses substantial challenges to traditional reductive evaluation approaches. In this paper we present an alternative methodological approach for studying how users' experiences with interactive products develop over time. The approach lies in the elicitation of rich qualitative insights in the form of experience narratives, combined with content-analytical approaches for the aggregation of idiosyncratic insights into generalized knowledge. We describe a tool designed for eliciting rich experience narratives retrospectively, and illustrate this tool by means of a study that inquired into how users' experiences with mobile phones change over the first 6 months of use. We use the insights of the study to validate and extend a framework of temporality proposed by Karapanos et al. (2009b).

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Understanding and designing for prolonged use is increasingly becoming a critical issue in the consumer electronics industry as evidenced by a number of recent trends (c.f. Karapanos et al., 2009b). First, legislation and competition within the consumer electronics industry has resulted in an increase in the time-span of product warranties, which lead to an alarming increase in number of products being returned on the basis of failing to satisfy users' true needs (Den Ouden, 2006). Secondly, products are increasingly becoming service-centered. Often, products are being sold for lower prices and revenues are mainly coming from the supported service. Thus, the measure of success for a product shifts from initial purchase to establishing prolonged use.

Critical to designing for prolonged use is our ability to measure the dynamics of users' experiences over time. Traditional approaches to measuring the dynamics of experience over time rely on validated measurement and structural models applied in different phases in the adoption of a product (e.g. Venkatesh and Davis, 2000). Such reductive approaches to measurement enable a systematic assessment of the changes in users' perceptions as well as in the relationships between different latent constructs. However, as it will be argued in this paper, they share a number of limitations in the measurement of the dynamics of experience over time.

First, these approaches assume that the relevant latent constructs (i.e. measurement model) remain constant, but that their perceived value and relative dominance in their effects on final outcome measures (i.e. structural model), such as overall evaluations of the *beauty* or *goodness* of an interactive product (Hassenzahl, 2004), change over time. Karapanos et al. (2009b) illustrated, however, how constructs such as *emotional attachment* and *daily rituals* become salient only after prolonged use. Consequently, this might challenge the content validity of the measurement model as relevant constructs that become salient only in prolonged use may be absent in measurement models that have only been validated during initial use. Second, by imposing a set of predefined measurement items, respondents are forced to aggregate the richness of their experience with a product, in a single attitudinal response. Especially in cases where respondents fail to interpret the relevance of a given statement in their own context, for instance when a construct has ceased to be relevant over prolonged use, the validity of the elicited data may be questioned (Larsen et al., 2008). Third, being quantitative in nature, such approaches provide limited insight into the exact reasons for the observed changes in users' experiences.

In this paper we propose an alternative approach to the measurement of the dynamics of users' experiences with interactive products. The approach relies on the elicitation of idiosyncratic self-reports of one's experiences, so-called *experience narratives*. Through a content analysis of the experience narratives, the researcher may acquire aggregated knowledge such as, for instance, changes in the relative dominance of different product qualities

* Corresponding author.
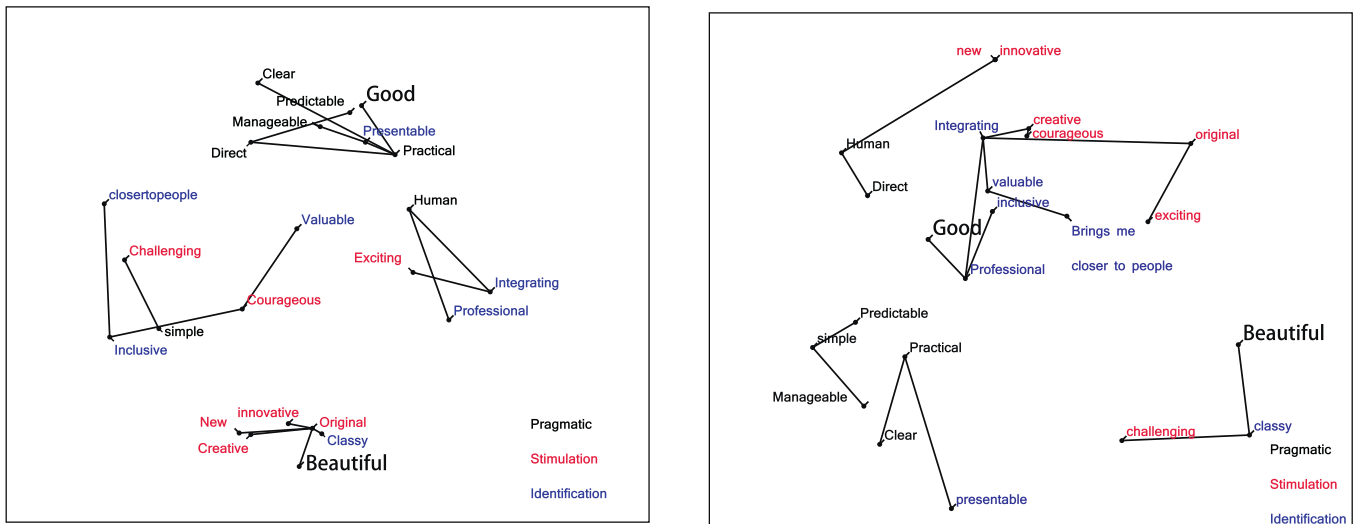 E-mail address: E.Karapanos@tue.nl (E. Karapanos).

Fig. 1. 2D projections of distances between quality attributes, beauty and goodness, representing users' perceptions during the 1st week (left) and 4th week of use. Items lying closely together are highly correlated. Items connected through a line belong to the same cluster.

over time. We illustrate iScale, a survey tool designed with the aim of minimizing retrospection bias in the elicitation of experience narratives, and employ this tool in a study that assessed the changes in users' experiences with mobile phones.

2. Measuring the dynamics of experience

This section reviews existing approaches to measuring the dynamics of user experience and introduces the methodological approach that we propose in this paper.

2.1. Reductive versus holistic approaches

Reductive approaches in the measurement of the dynamics of experience employ validated measurement and structural models across different phases in the adoption of a system. For instance, Venkatesh and Davis (2000) employed the Technology Acceptance Model (Davis et al., 1989) at three instances in the adoption of information systems at work settings: before the introduction of the system (inquiring into users' expectations), right after the introduction of the system, and 3 months after the introduction.

In the field of User Experience, we (Karapanos et al., 2008) employed Hassenzahl's (2004) model of pragmatic/hedonic quality, in exploring the differences (if any) in the ways that users evaluate interactive products during initial and prolonged use. Hassenzahl's (2004) model distinguishes between two kinds of quality perceptions: *pragmatic* and *hedonic*. Pragmatic quality, he argues, refers to the product's ability to support the achievement of behavioral goals (i.e. usefulness and ease-of-use). On the contrary, hedonic quality refers to the users' self; it relates to *stimulation*, the product's ability to stimulate and enable personal growth, and *identification*, the product's ability to address the need of expressing one's self through objects one owns.

Ten participants received an Interactive-TV set-top box containing a novel pointing device. We asked participants to evaluate the pointing device at two points in time, more specifically, during the first week and after 4 weeks of use, using the Attrakdiff₂ questionnaire (Hassenzahl, 2004). The relations between product qualities and overall evaluations were explored at these two points in time, 1st week and 4th week (see Fig. 1). We found that while in early interactions judgments of the overall goodness were primarily related to the perceived pragmatic quality of the product (i.e. utility

and ease-of-use), in prolonged use goodness related mostly to quality perceptions pertaining to identification (i.e. what the product expresses about its owner).

We argue that such reductive approaches are limited in a number of respects. An assumption inherent in these approaches is that the relevant latent constructs (i.e. measurement model) remain constant, but their perceived value and relative dominance (i.e. structural model) change over time. However, especially in developing fields such as user experience, substantial variations might occur over time even in what constructs are relevant to measure. Some constructs, for example, novelty, might cease to be relevant while others, such as daily rituals, personalization, and self-identity (see Karapanos et al., 2009b, that were not evident in studies of initial use might become critical for the long-term acceptance of a product.

Firstly, this might challenge the content validity of the measurement model as relevant latent constructs might be omitted. This issue has been repeatedly highlighted in technology acceptance research with a number of studies reporting limited predictive power of the Technology Acceptance Model (Davis et al., 1989), in some cases accounting for only 25% of the variance in the dependent variable (Gefen and Straub, 2000). Lee et al. (2003) reported that *"the majority of studies with lower variance explanations did not consider external variables other than original TAM variables"*.

Secondly, it may also lead to distorted data as participants might fail to interpret the essence of a question when the latent construct that the researchers attempt to measure (e.g. perceived novelty) ceases to be relevant over prolonged use. In such a case, participants may infer the meaning of a question from other features such as the language and the graphical layout of the questionnaire (see Schwarz, 2007). Larsen et al. (2008) reviewed a number of studies employing psychometric scales in the field of Information Systems. They found in the majority of studies that the semantic similarity between items was a significant predictor of participants' ratings with accounted variance of up to 63%. In such cases, they argued, participants are more likely to have employed *shallow processing*, that is, responding to surface features of the language rather than attaching personal relevance to the question (c.f. Sanford et al., 2006).

Lastly, such approaches provide rather limited insight into the exact reasons for changes in users' experiences. They may, for instance, reveal a shift in the dominance of perceived ease-of-use

and perceived usefulness on intention to use a product (e.g. Venkatesh and Davis, 2000), but provide limited insight to the exact experiences that contributed to such changes.

In contrast, holistic approaches attempt, through qualitative inquiry, to gain rich insights into users' idiosyncratic experiences. They attempt to inquire into the ways that design attributes interact with contextual details in particular contexts (Suri, 2002). One example is the elicitation of experience narratives, idiosyncratic self-reports on one's experiences with a product. These narratives provide rich insights into the experience and the context in which they take place. However, generalized knowledge can also be gained from these experience narratives. Such generalized knowledge consists of quantitative answers to questions like: how frequent is a certain kind of experience, what is the ratio of positive versus negative experiences and how does this compare to competitive products, how does the dominance of different product qualities fluctuate over time and what should designers improve to motivate prolonged use? This approach has a number of benefits over traditional reductive approaches. First, it provides a more complete coverage of users' experiences over time as participants are free to report any experience that is personally meaningful to them. Reduction and quantification of users' experiences only occurs in the analysis stage performed by the researcher. Second, it elicits data that are personally meaningful to the participant in contrast to reductive approaches which force participants to summarize their experiences according to imposed dimensions. Third, it allows linking generalized, quantifiable knowledge to rich and idiosyncratic insights. Both are important for design: idiosyncratic information may inspire design solutions, generalized knowledge may direct the focus of design.

An example is a study by Karapanos et al. (2009b). Following our initial exploratory insights into the dynamics of users' experiences we posed the following questions: what causes these changes? Can we describe the adoption of a product in terms of distinct phases? And which product qualities dominate in each of these phases? We followed six individuals through an actual purchase of an Apple iPhone. For a period of 4 weeks, participants were asked, during each day, to mentally reconstruct (Kahneman et al., 2004) their activities and experiences that somehow related to the specific product, and report the three most impactful experiences of the day in the form of one-paragraph essays, called *experience narratives*. A total of 482 experience narratives were elicited and submitted to a qualitative content analysis (Hsieh and Shannon, 2005). Narratives were coded with respect to the dominant product quality in each experience. The distribution of different experiences over time led to the realization of three overall phases in the adoption of the product: *Orientation, Incorporation* and *Identification* (see Fig. 2). In this paper we describe a study that attempts to validate and extend this framework.

Orientation relates to participants' initial experiences that are pervaded by feelings of excitement and frustration as they are confronted with *novel* features and *learnability* problems. Incorporation relates to reports of how the product is becoming meaningful in participants' daily lives as well as the realization of *usability* and *usefulness* problems that pertain over long-term use. Finally, identification relates to participants' *personal* experiences as the product plays a part in *daily rituals*, as well as to social experiences, as the product allows for *self-expression* and *creating a sense of community*.

2.2. Longitudinal versus retrospective approaches

A number of different longitudinal approaches exist for the elicitation of self-reports on one's experiences with a product.
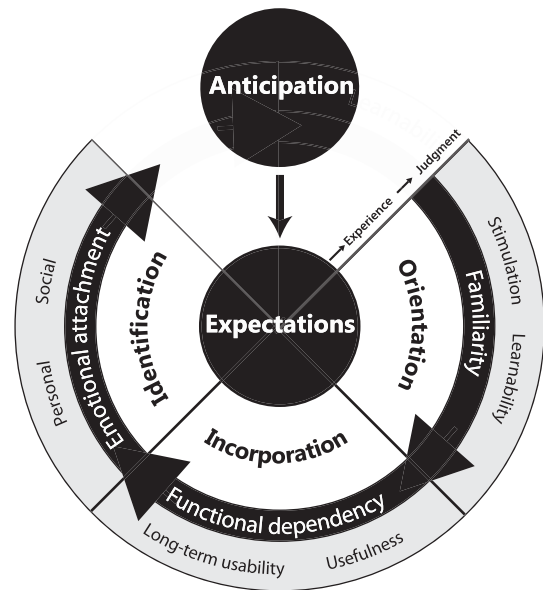

Fig. 2. A framework of *Temporality of experience* proposed in Karapanos et al. (2009b). Temporality consists of three main forces, an increasing *familiarity*, *functional dependency* and *emotional attachment*, all responsible for shifting users' experience across three phases: *orientation, incorporation* and *identification*. In each phase, different product qualities are salient.

Event-contingent diaries (see Bolger et al., 2003) ask participants to report experiences that they consider substantial enough at the moment of their occurrence. The Experience Sampling Method (Hektner et al., 2007) prompts participants to report on their current experience at random or computationally estimated times, for example, through context and activity detection algorithms. The Day Reconstruction Method (Kahneman et al., 2004), motivated by the need for more unobtrusive techniques, asks the participant at the end of each day to reconstruct, in a temporal sequence, the activities that took place during the day. This procedure has been found to assist participants in recalling accurately their daily experiences, while imposing less burden to participants being an offline method. In Karapanos et al. (2009b), we employed the Day Reconstruction Method in a four-week diary study. Each day participants were being asked to mentally reconstruct their daily experiences that somehow related to the product of interest, and to report the three most personally impactful experiences of the day.

Such longitudinal approaches may provide rich insights into users' experiences in different contexts. However, they are increasingly laborious when one needs to generalize over large populations of users and products, and inquire into long periods of time. Motivated by this observation, we developed *iScale*,[1] a survey tool that aims at eliciting users' most impactful experiences with products *retrospectively* (Karapanos et al., 2009a). iScale asks participants to narrate their most impactful experiences and to provide an estimation of when each experience took place, thus resulting in a temporal structure. Using *iScale*, one may inquire into the whole lifespan of a product in a single survey contact, in other words, participants may be asked to reconstruct their experiences from the moment of purchase till the present.

As *iScale* is a retrospective technique, one may wonder about the degree to which these experience reports are biased or incomplete. However, we argue that the veridicality of one's remembered experience is of minimal importance (Karapanos et al., 2009a), as

---

[1] A video demo of the tool is available at: http://ekarapanos.com/iscale.
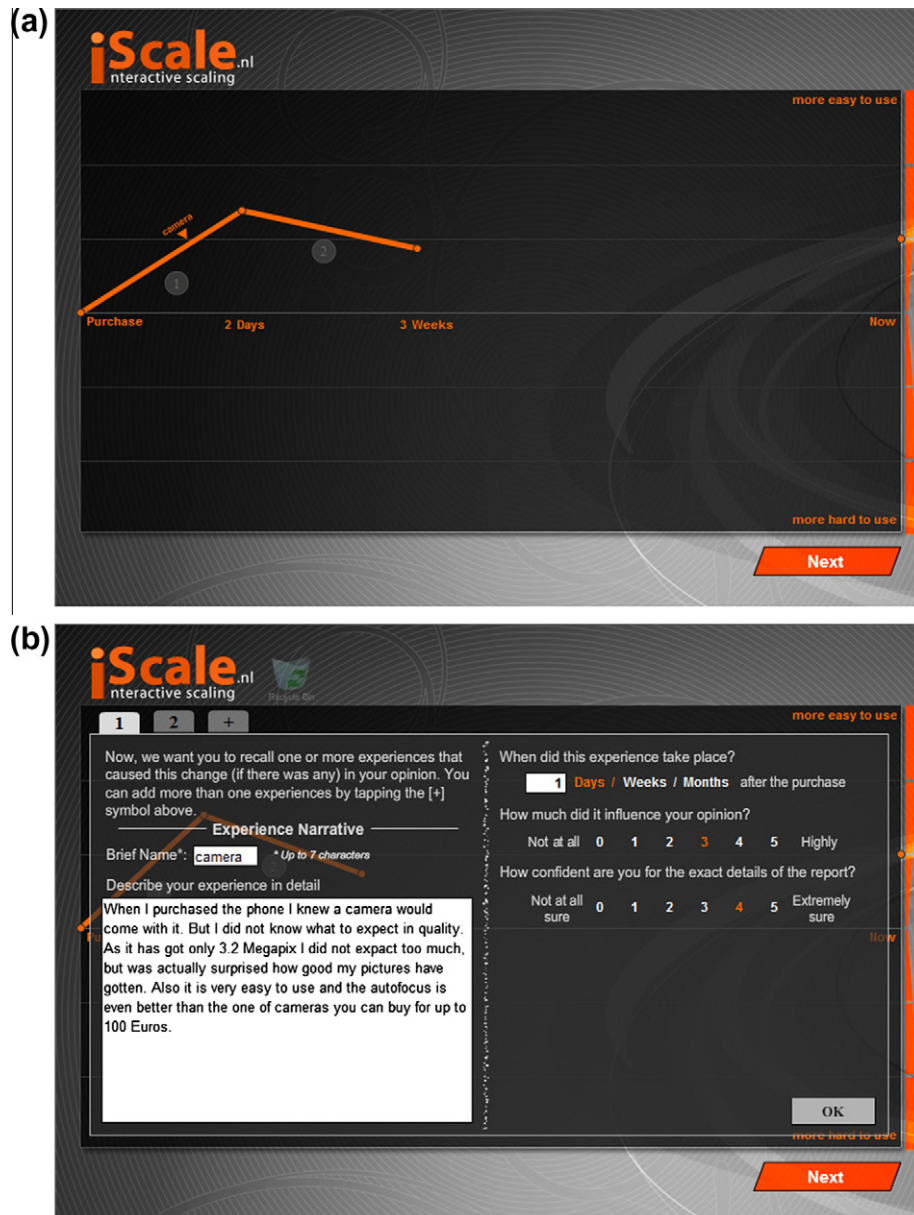
Fig. 3. iScale. A survey tool that elicits experience narratives retrospectively. (a) Users' sketch patterns that represent their perception of a given product quality. (b) Users report one or more experiences for each sketched line segment.

these memories (and not the actual experiences) (1) will guide future behavior of the individual and (2) will be communicated to others (see also Norman, 2009). In other words, it may not matter how a product was experienced in a given situation, but what individuals remember from this experience.

Although the validity of remembered experiences may not be crucial, their reliability is (Karapanos et al., 2009a). It seems at least desirable that participants would report their experiences consistently over multiple trials. If recall is random in the sense that different experiences are perceived to be important at different recalls, then the importance of such elicited reports may be questioned. In Karapanos et al. (2009a), we described the theoretical grounding and empirical testing of two different versions of iScale that were motivated by two distinct theories of how people reconstruct emotional experiences from memory. The iScale tool uses sketching in imposing a certain process in the reconstruction of experiences. It was found that these processes have a substantial impact on the number, the richness and the test–retest reliability

of recalled information when compared to a control condition, a free recall without employing any form of sketching (Karapanos et al., 2009a).

Participants using *iScale* are asked to sketch how their opinion on a given product quality (e.g. the perceived usability of a product) has changed as a course of time. In essence, participants sketch linear segments with a certain slope (which denotes the change in the value of the reported quality) and length (which denotes the

Table 1
The product qualities that participants reported on, along with definitions and word items.

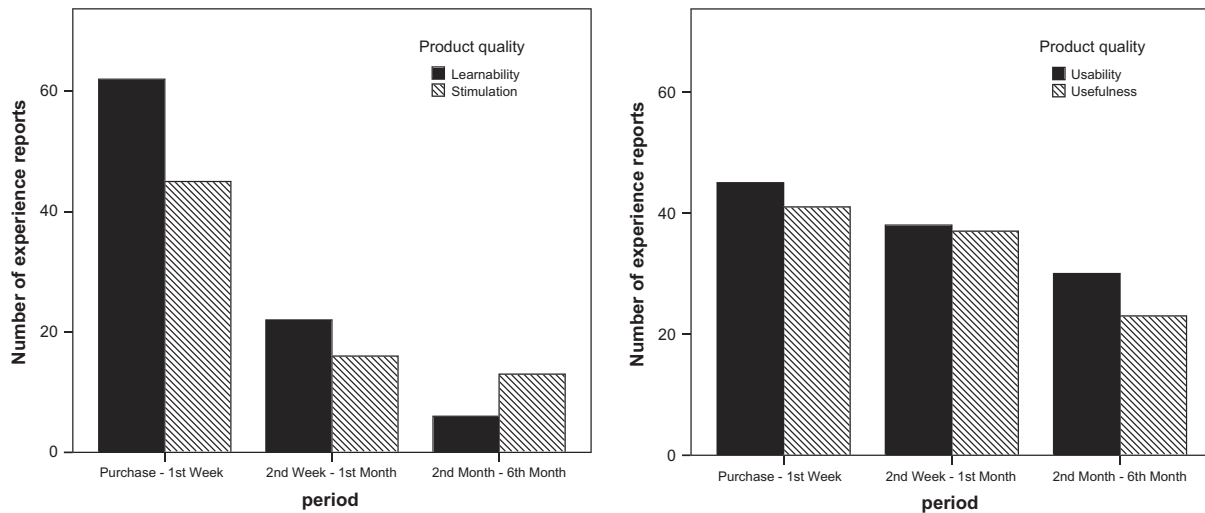| Name	Definition	Word
Ease-of-use	The ability of a product to provide the functions in an easy and efficient way	Easy to use, simple, clear
Innovativeness	The ability of a product to excite the user through its novelty	Innovative, exciting, creative

Fig. 4. Number of narratives in a certain period relating to learnability, stimulation, long-term usability and usefulness as a function of time of ownership.

temporal span of the reported period) (see Fig. 3a). Participants may then associate each line segment with one or more experiences that are believed to have induced the given change in their perception of the product quality (see Fig. 3b).

Two kinds of information are provided by *iScale*: (a) a list of experience narratives that are considered most impactful by the participant, and (b) a sketched pattern of the participant's perceptions of a given product quality. In the remainder of the paper we describe a study that employed iScale in reconstructing participants' experiences with mobile phones and propose a methodological approach for the analysis of the dynamics of experience over time.

3. A study on the dynamics of users' experiences with mobile phones

The study aimed at validating the framework of temporality proposed in Karapanos et al. (2009b). While the initial study was restricted to the first month of use and one particular product, the Apple iPhone, the present paper aims at extending the findings to the first 6 months of using mobile phones. While we limited the focus of the study to one product category, we did not opt to control further the type of product.

3.1. Method

3.1.1. Participants and material
The study was conducted with 48 participants (17 Female), ranging from 18 to 28 years old (median = 23 years). They were all students at a technical university; 19 of them majored in management related disciplines, 16 in design, and 13 in natural sciences and engineering. All participants owned a mobile phone for no less than four and no more than 18 months (median: 10 months); 16 participants owned a smart phone. Each participant joined a 1-h session and was rewarded by 10 euros for their efforts.

3.1.2. Product qualities
Hassenzahl (2004) proposed that people perceive products along two primary dimensions: *pragmatic quality* which refers to the product's ability to support the achievement of do-goals such as making a telephone call, and *hedonic quality* which refers to the product's ability to support the achievement of be-goals such as being happy. In the present study, we used these dimensions to differentiate pragmatic from hedonic experiences. Participants were asked to recall their mobile phone related experiences on

and rate a change in two product qualities: *ease-of-use* and *innovativeness*. Ease-of-use reflected pragmatic experiences while innovativeness reflected hedonic experiences; we avoided using the original terms as participants might fail to understand them. Product qualities were defined through a brief description and three single word items derived from the Attrakdiff$_2$ scales (Hassenzahl, 2004) (see Table 1). These word items reflected the positive pole of the three Semantic Differential scales that displayed the highest loading to the respective latent construct of the Attrakdiff$_2$ questionnaire in a previous study (Karapanos et al., 2008).

3.1.3. Procedure
Participants used the iScale tool in sketching how their perceptions of their mobile phone, regarding these two product qualities, developed from the moment of purchase till the present time. While sketching, they reconstructed and reported the experiences that affected their opinion about the product. For each drawn line segment (see Fig. 3a), they were asked to add at least one experience that induced this change in their perception of the respective quality (see Fig. 3b). Participants also annotated the timeline of the iScale tool by defining the estimated time that had passed from the purchase of the product till the end of each respective period.

3.1.4. Assumptions
A number of tentative assumptions were formed on the basis of the study presented in Karapanos et al. (2009b). The majority of users' experiences with iPhone relating to the product qualities *stimulation* and *learnability* displayed a sharp decrease over time: 58% of all experiences relating to stimulation and 59% of experiences relating to learnability took place during the first week of use. We expected this finding to be apparent also in the case of a wider spectrum of products such as the ones used in the current study, and not to be tied to the nature of iPhone, the product used in the original study (Karapanos et al., 2009b). We further expected this pattern to continue beyond the first month of use, with only a small number of experiences relating to learnability or stimulation being reported after the first month. On the contrary, *long-term usability* and *usefulness* became the dominant source of satisfying and dissatisfying experiences over time. We expected this to hold also in the current study and extend beyond the first month of use. Note that we did not distinguish between satisfying and dissatisfying experiences relating to a given product quality as they both tend to display a similar pattern over time (Karapanos et al., 2009b).
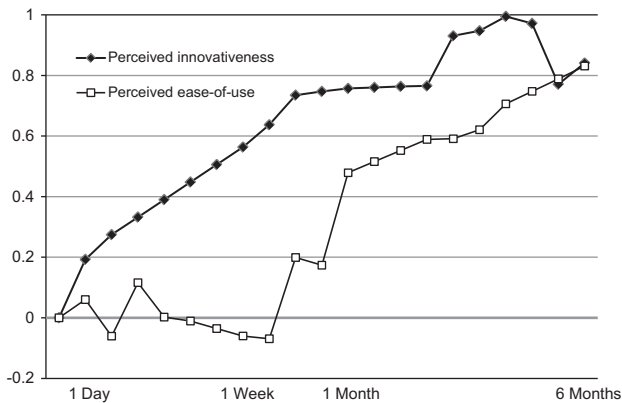
Fig. 5. Averaged across participants sketches for ease-of-use and innovativeness.

3.2. Results

A total of 398 experience narratives were elicited in the study. Each participant produced an average of 4.7 ± 1.4 experience narratives for each product quality. Due to the small number of experiences per participant and the limited variation across participants, an across-subjects analysis will be presented.

Surprisingly, 75% of all experience narratives related to the first month, while the median ownership time was 10 months. One would assume that recent experiences are more easily accessible in episodic memory and would thus be over-represented in a retrospective elicitation method (see Koriat et al., 2000). This hypothesis was not supported by the data. Instead, the temporal proximity of the experience to important milestones, such as the purchase of the product, seemed to affect participants' ability to recall or experience impactful events (Barsalou, 1988). Note, that this decrease in the number of recalled experiences over time was not affected by participants' motivation in reporting which is often apparent in longitudinal studies as recall took place retrospectively and at a single session; instead, we are more inclined to attribute this to participants' varying ability to experience and recall impactful events over time. Consequently, the mental representation of time is expected to be affected by the variable accessibility of events pertaining to different periods. This was indeed found in our analysis reported in Karapanos et al. (2009a), where the time-scale (i.e. *x*-axis in the iScale tool) was found to relate to actual time [2] through a power-law. In other words, participants had a tendency to use a substantial fraction of the *x*-axis of iScale to map their initial experiences. Based on these two insights, it was decided to restrict the analysis only to the first 6 months of usage; the remaining experience narratives were discarded. In addition, we chose to divide the time in a non-linear way. Specifically, Fig. 4 distinguishes three time intervals: (a) purchase-1st week, (b) 2nd week–1st month and (c) 2nd month–6th month. In the same vein, Fig. 5 represents participants' sampled perceptions of ease-of-use and innovativeness at 1st day–7th day, 2nd week–4th week and 2nd month–6th month.

3.2.1. Content-analysis of experience narratives

Experience narratives were submitted to a qualitative content analysis (Hsieh and Shannon, 2005) using an a-priori defined classification scheme imposed by the temporality framework (Karapanos et al., 2009b). The coding was performed by the first author and a small set (10%) of the data were re-coded by an external researcher (interrater aggreement: K = 0.73, Fleiss et al., 2003). We found experiences to tap into a wider spectrum of product qualities

and not only into the ones that participants were asked to report. Overall, we identified four categories out of the initial six categories of the temporality framework. These were: usefulness, long-term usability, learnability, and innovativeness. Social aspects of product use did not become apparent. Such experiences were rare also in the initial study of Karapanos et al. (2009b).

In support of prior work (Karapanos et al., 2008, 2009b), the dominance of learnability and stimulation experiences decreased over time (see Fig. 4). Out of all experiences relating to learnability and pertaining to the first month of use, 74% of them took place during the first week of use. This is even higher than the 59% of the original study (Karapanos et al., 2009b). An example of such an early learnability problem is as follows:

> "[learnability, 1 week] *I had to make a new contact list. This took a lot of my time. It wasn't possible to copy my contact list to my new phone. I had to type every contact separately*"

Only a small number of experiences, 7% (12) of all experiences relating to learnability, were estimated to have taken place after the first month of use. These mostly related to participants' reflection on their increasing familiarity with the product:

> "[learnability, 1 month] …*it took one month to find out how to activate the keypad lock because it was not indicated in the manual and it was not obvious on the phone*"

> "[learnability, 1 month] *After some weeks, using my phone went better and easier. I figured out how I could use the Menu properly and the time I spend on making a text message was shorter. So I was glad about the purchase again*"

In a similar vein, 75% of experiences relating to stimulation took place during the first week of use, as compared to the figure of 58% that was found in the original study. This difference between the two studies, which is apparent in both product qualities, might be an effect of the different products being studied, but also an effect of the different method, being longitudinal or retrospective. In the original study, participants were asked to report daily, over a period of one month, the three most impactful experiences. Contrary, the current study asked participants in a single session to consider the full time of ownership of the product and narrate the most impactful experiences throughout this whole period. In other words, the first study informs us about the relative dominance of experiences within a single day; the latter one informs us about the relative dominance of remembered experiences throughout the whole lifespan of a product. Thus, this difference between the two studies might pertain to the discrepancy between what users experience on a daily basis and what they remember after some time.

Next, while experiences relating to stimulation displayed a decreasing pattern, a considerable number of experiences (17%) took place after the first month. The following reasons for this finding were identified.

Firstly, the novelty of a feature was often not apparent in initial use, but only after participants used it in a real context, or appropriated it in different contexts:

> "[innovativeness, 4 months] *With this mobile I can take pictures of the faces and then edit them in a funny way, such as making them look like an insect or an alien, which is a feature of the mobile. It also has some funny frames in which I can put peoples faces have fun*"

Secondly, in certain cases participants had low expectations about the product, leading to feature discovery at a later stage in product adoption:

---

[2] Each node in the iScale graph is annotated by the exact time which differentiates the two aligning periods (see Fig. 3).

"[innovativeness, n/a] *When I first bought the mobile I did not know that it had web access, also because I did not care much about it…*"

Thirdly, after some time participants started reporting a decrease in perceived innovativeness due to new products coming out in the market or acquired by others in their social circles.

"[innovativeness, n/a] *After 2 months there were coming more and more gadgets on the market. Like iPhones and Blackberrys. The number of functions of those phones is increased strongly. So when I compare my phone to those phones, it seems to be not as innovative as I thought before. Having a mobile phone with a color screen, radio and music, is normal in today's life*"

As expected, users' experiences increasingly related to aspects of *usefulness* and *long-term usability*, and the value of the product was derived through its appropriation in different (and often unexpected) contexts. While the number of these experiences decreased over time, their relative dominance over experiences relating to stimulation and learnability increased.

"[usefulness, 2 months] *I found a new use for an option on my phone; you can leave a spoken message on it. So I use it for short messages to myself*", [long-term usability, 2 months] "*The button at the side of the phone allows easy access to the camera function…This is extremely convenient when we want to take photos while wearing gloves.*"

Similarly, users were often dissatisfied by features that did not sustain their value or through long-term usability flaws that became apparent after some usage.

"[usefulness, 2 months] *Although at the beginning I enjoyed taking pictures with my mobile phone, later when I saw that the quality is really low, i did not use it as much as I did before*"

"[usefulness, NA] *I thought I would get used to the menu but after such a long time I still do not like it. My phone remembers where I left the menu the last time, so when I open it, it starts there. But ending there once doesn't mean for me that I want to start there the next time. That's a pity and I can't adjust the settings of it*"

Overall, out of all experiences that took place in the first month of use ($n = 305$), we found a significantly higher percentage of these experiences to relate to stimulation (20% as opposed to 9% in the original study, χ^2, $p < .001$), while a significantly lower number of experiences related to usefulness (26% as opposed to 36%, χ^2, $p < .001$). One tentative explanation of this difference might be grounded in the retrospective experience elicitation method used in this study as compared to the longitudinal method of the original study. In other words, while users' reality might be dominated by "*dull*" experiences relating to the product's usefulness, they might have a bias for remembering the most emotionally aroused ones such as experiences relating to stimulation (Christianson, 1992).

3.2.2. Averaging sketched patterns across participants

Next, an overall pattern of the perception of the two qualities, i.e. ease-of-use and innovativeness, can be obtained by averaging participants' sketches across time (see Fig. 5). Participants' sketched patterns were sampled according to actual time (i.e. the reported day that each experience took place), to construct average graphs of how two product qualities, *ease-of-use* and *innovativeness*, changed over time. In this sampling, each participant's sketched pattern is transformed to an actual timeline through the use of the self-reported exact time for each node of the graph (see Fig. 3) and the perceived value (*y*-coordinate) is captured.

The resulting averaged pattern suggests that users' perception of the innovativeness of mobile phones increased during the first month and then remained approximately stable. On the contrary, users seemed to experience learnability problems mostly in the first week of use; after this period usability displayed a sharp increase over the course of the first month while this increase continued more moderately till the end of the studied 6-month period. While users continue reporting problems relating to the long-term usability of the product (see Fig. 4), these problems seem to have lower impact on users' perceptions of the product's ease-of-use. This was also observed in Karapanos et al. (2009b), whereas while long-term usability issues constituted one of the most frequent sources of negative experience over prolonged use, such issues were rated as less severe, when compared to problems pertaining to learnability.

4. Discussion and conclusion

In this paper we highlighted the limitations of traditional reductive approaches in the measurement of the dynamics of user experience over time. We argued that substantial variations might occur over time in the constructs that are relevant to measure, which creates at least two complications. Firstly, it challenges the content validity of the measurement model as relevant constructs may be omitted. Secondly, it may lead to unreliable data as participants may fail to interpret the relevance of a given item scale to their own context when the latent construct and its individual scale items cease to be relevant.

An alternative approach that relies on the elicitation of idiosyncratic experience narratives was proposed in this paper. We proposed that experience narratives may lead to two kinds of information. First, they provide idiosyncratic insights in a given experience within a situated context. Such information may provide a rich understanding of how design attributes interact with contextual details in given contexts (Suri, 2002) and may inspire design solutions (Hornbæk, 2008). Next, narratives may be submitted to a content analysis in order to extract generalized knowledge such as: what are the different types of experience and how do they develop over time? These two insights enable the researcher to shift between different levels of abstraction, from highly idiosyncratic and design-relevant information to generalized knowledge. This shift is essential as it enables the designer to perform rapid hypothesis testing. While experience narratives may provide insightful information one cannot generalize to different contexts and user populations. At the same time, designers may run the risk of focusing on interesting but rare experiences, what Kahneman et al. (1982) called the *availability* bias, a heuristic that people apply when judging the relevance or probability of occurrence of a specific event where more easily retrievable, or available, events in an individual's memory are given higher weight. In other words, unique experiences, or ones that strongly deviate from users' typical responses to the product, will evidently be more salient in designers' memory and may be over-represented in design.

We further argued for retrospective techniques as a cost-effective alternative to longitudinal field studies in the elicitation of experience narratives. We presented iScale, a survey tool that assists participants in recalling their most impactful experiences with a product while sketching how their perception on a given quality developed over time. We argued that while such memories may differ substantially from in situ longitudinal data, these memories are more important than the actual experiences as the memories and not the actual experiences eventually affect an individual's attitude towards a product, or what he or she communicates to others.

One potential criticism of the proposed approach is that, given that such retrospections may span long periods of time (e.g. 6 months or 1 year) and that we expect time to affect users' memories, designers' goals become a moving target depending on the time that has elapsed from moment in which the experience took place till the time of data collection. From a design perspective, we disagree. We believe that this bias does not only exist in measurement, but instead, that it is ubiquitous in life, and that designers need to account for such biases. Consider, for instance, the process by which a user assesses an ongoing experience. Such assessments are typically conducted in comparison to reference points: experiences that took place in the past. The 'past' may range from just a few moments ago to many years ago. Such-reference-memories will anyhow be biased representations of the actual experiences and the length of retrospection is likely to affect the elicited data. As designers, we have no control for this. We thus argue that design is always a moving target. In other words, completeness (i.e. do we capture all possible ways in which users' might respond to a product?) matter more than validity.

Still, designers' are not always interested in users' memories. Often, the actual, and not the remembered, experiences should be at the forefront. Consider, for instance, the case where we might want to know the reasons that underly non-responsible driving behavior. Memories offer little understanding as to what motivates such behaviors. Retrospective techniques are not aimed to replace longitudinal field studies and in-situ methods. Instead, we propose that retrospective techniques may be a viable alternative to longitudinal studies when memories are placed at higher importance than actuality.

We illustrated this methodological approach through a study that aimed at validating and extending a framework of temporality proposed by Karapanos et al. (2009b). Overall, we found limited differences in the way users' experiences develop over time, between the initial study that considered only one product and a limited amount of time and the current study that extended to more products and beyond the first month of use. We found, however, that the majority of users' experiences with mobile phones (75%) to pertain to the first month of use. This result might hint the commodification phase of this particular product category, the moment at which these products stop providing users personally impactful experiences and become a commodity (Silverstone and Haddon, 1996).

One of the limitations of the proposed approach is its labor-intensive nature as a large amount of experience narratives need to be content-analyzed in the development of theory of the dynamics of users' experiences over time. Our current work explores semi-automated approaches for the analysis of experience narratives. Such approaches combine traditional qualitative coding procedures (Strauss and Corbin, 1998) with computational techniques for assessing the semantic similarity between documents (Salton et al., 1975). As a result, they provide the ability to automate the analysis process, as concepts that are coded in the analysis of one narrative are automatically indexed and queried in all narratives, while maintaining a level of interpretation that is necessary in qualitative analysis and is not supported through traditional semantic analysis techniques (e.g. Deerwester et al., 1990). Next, an interesting question not currently addressed is the identification of causal relations between identified concepts within and across different experiences, and how these develop over time. Future work will also address this aspect.

References

Barsalou, L., 1988. The content and organization of autobiographical memories. In: Neisser, U., Winograd, E. (Eds.), Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory. Cambridge University Press, pp. 193–243 (Chapter 8).

Bolger, N., Davis, A., Rafaeli, E., 2003. Diary methods: capturing life as it is lived. Annual Review of Psychology 54 (1), 579–616.

Christianson, S., 1992. The Handbook of Emotion and Memory: Research and Theory. Lawrence Erlbaum.

Davis, F.D., Bagozzi, R.P., Warshaw, P.R., 1989. User acceptance of computer technology: a comparison of two theoretical models. Management Science 35 (8), 982–1003.

Deerwester, S., Dumais, S., Furnas, G., Landauer, T., Harshman, R., 1990. Indexing by latent semantic analysis. Journal of the American Society for Information Science 41 (6), 391–407.

Den Ouden, E., 2006. Development of a design analysis model for consumer complaints. Ph.D. Thesis, Eindhoven University of Technology.

Fleiss, J.L., Levin, B., Paik, M.C., 2003. Statistical Methods for Rates and Proportions. Wiley-Interscience.

Gefen, D., Straub, D., 2000. The relative importance of perceived ease of use in IS adoption: a study of e-commerce adoption. Journal of the Association for Information Systems 1 (8), 1–28.

Hassenzahl, M., 2004. The interplay of beauty, goodness, and usability in interactive products. Human–Computer Interaction 19 (4), 319–349.

Hektner, J.M., Schmidt, J.A., Csikszentmihalyi, M., 2007. Experience Sampling Method: Measuring the Quality of Everyday Life. Sage Publications Inc.

Hornbæk, K., 2008. Usability evaluation as idea generation. In: Law, E.L.-C., Hvannberg, E.T., Cockton, G. (Eds.), Maturing Usability. Springer, pp. 267–286.

Hsieh, H.F., Shannon, S.E., 2005. Three approaches to qualitative content analysis. Qualitative Health Research 15 (9), 1277–1288.

Kahneman, D., Slovic, P., Tversky, A., 1982. Judgment under Uncertainty: Heuristics and Biases. Cambridge University Press.

Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., Stone, A.A., 2004. A survey method for characterizing daily life experience: the day reconstruction method. Science 306 (5702), 1776–1780.

Karapanos, E., Hassenzahl, M., Martens, J.-B., 2008. User experience over time. In: CHI'08 Extended Abstracts on Human Factors in Computing Systems. ACM, Florence, Italy, pp. 3561–3566.

Karapanos, E., Martens, J.-B., Hassenzahl, M., 2009a. Reconstructing experiences through sketching. <http://arxiv.org/abs/0912.5343>.

Karapanos, E., Zimmerman, J., Forlizzi, J., Martens, J.-B., 2009b. User experience over time: an initial framework. In: CHI'09: Proceedings of the 27th International Conference on Human Factors in Computing Systems. ACM, pp. 729–738.

Koriat, A., Goldsmith, M., Pansky, A., 2000. Toward a psychology of memory accuracy. Annual Review of Psychology 51 (1), 481–537.

Larsen, K., Nevo, D., Rich, E., 2008. Exploring the semantic validity of questionnaire scales. In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences, p. 440.

Lee, Y., Kozar, K., Larsen, K., 2003. The technology acceptance model: past, present, and future. Communications of the Association for Information Systems 12 (1), 752–780.

Norman, D., 2009. THE WAY I SEE IT. Memory is more important than actuality. Interactions 16 (2), 24–26.

Salton, G., Wong, A., Yang, C., 1975. A vector space model for automatic indexing. Communications of the ACM 18 (11), 613–620.

Sanford, A., Sanford, A., Molle, J., Emmott, C., 2006. Shallow processing and attention capture in written and spoken discourse. Discourse Processes 42 (2), 109–130.

Schwarz, N., 2007. Cognitive aspects of survey methodology. Applied Cognitive Psychology 21 (2), 277–287.

Silverstone, R., Haddon, L., 1996. Design and the domestication of information and communication technologies: technical change and everyday life. In: Silverstone, R., Mansell, R. (Eds.), Communication by Design: The Politics of Information and Communication Technologies. Oxford University Press, pp. 44–74.

Strauss, A.L., Corbin, J.M., 1998. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Sage Publications Inc.

Suri, J., 2002. Designing experience: whether to measure pleasure. In: Green, W., Jordan, P. (Eds.), Pleasure with Products: Beyond Usability. Taylor & Francis, London, pp. 161–174.

Venkatesh, V., Davis, F., 2000. A theoretical extension of the technology acceptance model: four longitudinal field studies. Management Science 46 (2), 186–204.