Modelling user experience – An agenda for research and practice

1. Introduction

1.1. Motivation and rationale

A major area of research is how a positive ‘user experience’ – or interaction experience – of the use of digital artefacts (e.g. web sites, virtual worlds and personal digital assistants) can be promoted (Hassenzahl and Tractinsky, 2006). This experience does not only include usability, but also other cognitive, socio-cognitive and affective aspects of users’ experience in their interaction with artefacts, such as users’ enjoyment, aesthetic experience, desire to repeat use, positive decision to use a digital artefact and enhanced mental models. Research in this area is timely because we are approaching the ‘loyalty decade’, where interaction experience will become the main success factor (Nielsen, 2008). Hence, the success of digital artefacts is to a large extent positively influenced by the extent to which they promote a high-quality experience in their users.

Furthermore, user experience (UX) manifests as quality in design, in interaction and in value, with diverse measures from many methods and instruments. One of the challenges related to UX is how to select appropriate measures to address the particularities of an evaluation context. The necessity and utility of UX measures is apparent, because such measures enable professionals to benchmark competitive design artefacts and to select appropriate design options. However, both the construct validity and predictive power of some UX measures are of particular concern. Consequently, modelling users’ experience – as a basis for producing design guidance – is especially important.

Generally speaking, two types of model are distinguished in the behavioural sciences: whereas measurement models are used to measure the constructs in a particular domain, structural models are used to establish (causal) relations between constructs (Edwards and Bagozzi, 2000). These two types of model are indispensable to advance progress in a number of disciplines. First, sound measures need to be established with desirable properties (e.g. reliability, validity and sensitivity) to provide a sound basis for measuring UX. Second, explanatory or predictive structural models need to be developed – linking antecedents through behaviour to consequences – for the purpose of understanding, predicting and reasoning about processes of UX to inform system design. Irrespective of whether strict formal measurement paradigms are brought to bear on traditional human–computer interaction (HCI) phenomena like usability or emerging ones like UX, it is the persuasiveness of empirical evidence that is ultimately the test of its worth.

Despite some visible progress (e.g. van Schaik and Ling, 2008), a number of issues pertaining to measurement models and structural models of UX remain to be resolved and become relevant items of an agenda as a stimulus for further work. Some of these issues are addressed, although to various extents, by the five papers in this special issue on modelling UX. In the ensuing text each of these issues is presented together with practical implications and research questions it engenders. It is important to note that a review of measurement models and structural models of UX is beyond the scope of this publication.

1.2. A basic illustration of UX modelling

To set the scene for a more elaborate discussion of issues for research and practice, a basic illustration of UX modelling is presented, using a simple (in terms of complexity) and simplified (in terms of representation) model of UX, based on Hassenzahl (2004) and van Schaik and Ling (2008). The purpose of this exercise is not to present a perfect model (if this would be possible in the first place) of UX, but to demonstrate the main features of a measurement model and a structural model. The main UX constructs are a user’s perceived hedonic quality (pleasure–producing product qualities), pragmatic quality (user-perceived usability), beauty (aesthetics) and goodness (overall product quality).

The measurement model (Fig. 1a) represents these four constructs as latent variables that are measured using manifest variables (indicators). Both perceptions have four manifest variables each (PQ1–PQ4 and HQ1–HQ4) and both evaluations have one manifest variable (B1 and G1) – indicated using single-headed arrows. Data on the manifest variables are collected using questionnaire items, all with a 7-point rating scale. The measurement model allows latent variables to be correlated (indicated using double-headed arrows). The structural model (Fig. 1b) states the cause–and-effect relations (indicated using single-headed arrows) between UX constructs as latent variables and other variables. (In this simple example there is only one variable; this is usability – as a design characteristic, e.g. in terms of screen layout – of the artefact that is used.) According to the model, usability has a positive effect on pragmatic quality, but not on hedonic quality. Both pragmatic- and hedonic quality have a positive effect on goodness, but only hedonic quality has a positive effect on beauty. Once data have been collected, they are analysed to test the models, using techniques of ordinary least squares (see e.g. Tabachnick and Fidell, 2007), partial least squares (see e.g. Esposito Vinzi et al., 2010) or (covariance-based) structural equation modelling (see e.g. Kline, 2005). To start, the measurement model is tested. If the measurement model fits well with the data (and this is a prerequisite) then the structural model is tested.

2. Measurement models

The following conceptual and practical issues germane to UX measurement models are discussed: intertwined relationships...
between UX and usability, measurability of UX attributes, correlations between subjective and objective measures, and aggregated measures.

2.1. Intertwined relationships between UX and usability

Logically speaking, UX measures should be derived from a clear definition of UX. The formal definition of UX issued by ISO 9241-210 (2010, clause 2.15) – “A person’s perceptions and responses that result from the use and/or anticipated use of a product, system or service” – suggests that UX can be measured in a way similar to the behavioural and attitudinal metrics of usability (i.e. users’ performance and satisfaction). Indeed, different attempts have been undertaken to demarcate or even dismiss the boundary between usability and UX at the conceptual as well as operational level (e.g. Law et al., 2008). Two major stances on this issue are: first, usability is subsumed by UX; second, UX is an elaborated form of satisfaction – one of the three usability metrics.

Goals and tasks are employed to differentiate usability from UX – pragmatic/goals are associated with usability and hedonic/be goals are with UX, and the former is tagged as task-oriented and the latter is non-task-related (e.g. Bevan, 2009). Given the situatedness of goals that couple with contextual changes and the malleability of the notion of task (Draper, 1993; Lindgaard and Chattratchart, 2007; Wood, 1986), applying them as differentiators between usability and UX is questionable. For instance, in a game-based learning environment that amalgamates a pragmatic goal with a hedonic one, a particular ‘gamer’ dynamically updates his learning/gaming objective contingent on the nature of an interactive game activity. Such an activity can be an entertainment for one person and a chore for another. With research on demarcating UX still in progress, current methods, techniques and tools for evaluating and measuring UX are largely drawn from the usability tradition (cf. Roto et al., 2009). For instance, Tullis and Albert (2008) entitle their book “Measuring the user experience” (Note: the article ‘the’ may suggest the book focuses on a specific aspect of UX or it has no particular implication) and supplement it with a secondary title “collecting, analyzing and presenting usability metrics”, which seems to imply that measures of UX and usability cannot be differentiated.

2.1.1. Practical implications

A series of research activities for understanding, scoping and defining UX have been undertaken in recent years (see the review in Law et al. (2009)), resulting in a broad but yet unconsolidated body of knowledge of UX. Apparently, the basic issue about the distinction between UX and usability requires a deeper and more systematic conceptual analysis. Identification of the uniqueness of UX does not imply abandoning the traditional usability approaches, which should actually serve as the base for incorporating some new requirements of UX.

In accord with the common understanding of UX as subjective, dynamic and context-dependent (Law et al., 2009), UX measurement should essentially be self-reported, trajectory-based and adaptive. Traditional techniques such as questionnaire, interview, and think-aloud remain important for capturing self-reported data. With the advent of open-source, multimedia social software such as blogs and video-wiki (e.g. Law and Nguyen-Ngoc, 2008), it has become increasingly practical to capture as well as share experience over a range of timeframes and contexts. However, what could be impractical are the resources required to analyse a huge body of rich experiential data that would result from such an approach. Further, a trajectory-based approach implies measuring various aspects of UX both in different contexts and at different points of time: user-expectation (imaginary UX), user-affect (momentary UX) and user-emotion (long-term UX) (cf. Russell, 1980). This allows a clear picture of how UX changes over time. Supplementing with some qualitative approaches, reasons for the changes can also be derived (Karapanos et al., this volume). To estimate an overall UX score, when needed, different weights should be assigned to the measures taken at different points of time, especially user-expectation and user-affect dynamically evolve with the actual usage of the product over time. A challenge lies in developing effective algorithms to compute dynamic weights sensitive to changes in context.

2.1.2. Research questions

How are UX measurement models distinct from usability models, especially with respect to the temporal aspect – learnability in the case of usability and malleability of emotion in the case of UX? How can new information and communication technologies (ICT) be used to capture and document UX flexibly and economically? Can the resource-demanding process of analysing UX data be streamlined and automated, and, if so, how?

2.2. Measurability of UX attributes

Much evaluation work in HCI aspires to its scientific philosophy and adheres to the measurement paradigm, in spite of controversies over the measurability of some psycho-social constructs as mundane as beauty, happiness, frustration, and pain. While some HCI researchers and practitioners strongly advocate the necessity and utility of measurements (e.g. Sauro, 2006), some others are ambivalent about the role of numerical values in our (deep) understanding of complex interactions between humans and machines.
Obviously, from a practical (design) perspective, approximate measurements would be sufficient if they can help improve system design. Basically, one could measure anything in any arbitrary way, but the compelling concern is whether the measure is meaningful and valid to reflect the state of nature of the object in question. Some in HCI refute measurement, based on their defiance of reductionism, arguing that human experiences and feelings (or embodied interactions) should not and cannot be reduced to numbers. Such an argument reflects the decades-long dispute between the cognitivist and phenomenological approaches (Winograd and Flores, 1987; Dourish, 2001). Specifically, Boehner et al. (2007) discuss the measurability of emotion – the key construct of UX – with reference to the cognitive rationalistic model as opposed to the socio-constructive interactionist framework. In fact, the situated cognition movement dated back to the late-1980s has already instigated the concern about the excessive rendering of conceptual entities like cognition and emotion into experimental variables that are amenable to being measured, modelled and formalized.

Whilst we do not endorse the stance that all kinds of affective and emotional state can be reduced to numeric or graphical representations, meaningful and valid measures of these constructs are deemed necessary and useful for comparing as well as predicting the quality of the artefact contributing to the experience being measured. Echoing the recommendation of Gray and Salzman (1998), a multi-method multi-operation measurement approach should be adopted. Triangulation of mixed measures entails carefully crafted research protocols as well as in situ practices, and the related process of data collection and analysis can be very costly. Hence, it may not be feasible to implement this approach when resources are limited.

### 2.2.2. Research question

What properties should an instrument have that allows it to effectively measure individual experience as well as co-experience evoked in human–technology interaction and technology-mediated social interactions within the same time frame, either physically distributed or co-present?

### 2.3. The use of and relationship between subjective and objective measures

There remains a long-established argument in HCI over whether objective or subjective measurements should be used, especially for non-performance, experience-based constructs such as emotion (e.g. Spagnolli et al., 2003; Wilson and Sasse, 2004). The dispute lies not only in which of the two types of measure is more appropriate but also in whether and how they are related and under which conditions.

Earlier research studies show that objective measures such as time-on-task and completion rate do not necessarily correlate with subjective measures such as users’ preference and perceived satisfaction (e.g. Frøkjær et al., 2000; Kissel, 1995; Nielsen and Levy, 1994). Results of the meta-analysis conducted by Hornbæk and Law (2007) show that effectiveness, efficiency and satisfaction (i.e. the three prototypical usability measures) are correlated at a low to medium level. Among others, factors influencing the correlations include the complexity of measures and duration of use.

**Fig. 1 (continued)**
Sauro and Lewis (2009), in contrast, demonstrate stronger correlations between usability metrics when aggregating the measures at the task level.

Boehner et al. (2007) also discuss the issue of whether objective or subjective measures of emotion should be used, however, not in terms of their correlation. Instead, they argue for the superiority of self-reported measures as opposed to physiological ones. Psychophysiological measures have increasingly been employed in evaluating UX (e.g. Ganglbauer et al., 2009; Mandryk and Atkins, 2007; Shami et al., 2008), because of their purported power to capture experience as a stream of consciousness without interrupting a particular user’s experiential activity. These measures manifest as seemingly precise numbers and graphs, supporting their scientific flavour and appeal. However, without systematic calibrations, these data are rendered meaningless. Hence, it is a common practice to triangulate psychophysiological measures with self-reported data. Such a mixed measures approach is adopted by Nacke et al. (this volume), who demonstrate some significant correlations between certain facial muscle activity (EMG) and subjective responses to questions gauging several dimensions of gameplay experience.

Hedonic quality was empirically found to be a determinant of users’ evaluation of beauty, but both hedonic quality and pragmatic quality were determinants of goodness (users’ overall evaluation of product quality; Hassenzahl, 2004). However, inconsistent findings about the relation between usability and beauty are no less confusing than those between preference and performance (Nielsen and Levy, 1994); the sources for (a lack of) correlations are yet unclear (cf. for a brief review on this issue, Law and Hornbæk, 2007). Inconsistencies uncovered by studies investigating the correlation between usability and beauty could be experimental artefacts (cf. Monk’s (2004) fixed effect fallacy and the problem of using a single-item questionnaire), and could also be psychological phenomena such as halo effect, social desirability and calculative judgments based on economic as well as social value (Hartmann et al., 2008).

2.3.1. Practical implications

The utility of the subjective–objective distinction depends much on the intended context of use and different weights are assigned to specific measures. A balanced focus on both types of measure may help improve both UX and performance (Hornbæk, 2006). Furthermore, patterns in correlations within different sets of usability and UX measures can indicate the validity of measuring tools as well as the validity of underlying measurement models. Systematic analysis of these patterns in the form of a large-scale meta-analysis will be helpful to understand the relationships between subjective and objective measures on the one hand and between usability and UX on the other hand.

2.3.2. Research questions

Which theoretical frameworks can explicate the relations between usability and UX measures and support the prediction under which conditions should they be (not) correlated? Although perceived aesthetics is measured subjectively (Lavie and Tractinsky, 2004), it seems that objective measurement with quantifiable parameters has not been achieved yet (cf. the interesting neurophysiological study of Kawabata and Zeki (2003)). How can this challenge be tackled? Can a ratio of (weighted) subjectively/objectively measured aesthetic values be computed as a specific UX metric (cf. the proposed subjectively perceived/objectively assessed duration as a usability metric; Czerwinski et al., 2001), and, if so, how?

2.4. Aggregated measures

Assuming that a single composite standardized measure of usability is more useful and usable than a set of numbers and tables for managers or other organizational decision-makers to position their product, industrial researchers have recently attempted to combine usability measures with the argument that such an aggregated measure can convey essentially the same information. However, the validity of this argument has been challenged (see the critiques on the related work of McGee (2004) and Sauro and Kindlund (2005) in Hornbæk and Law, 2007). The more recent work of Sauro and Lewis (2009) shows higher correlations among the usability metrics as compared to those reported by Hornbæk and Law (2007). The difference in findings between the two studies is explained by the different levels of data aggregation (task or user) that influence the magnitude of correlation among usability metrics. Whilst Sauro and Lewis (2009) acknowledge that the main limitation of a summary score is information loss, they still argue for its necessity in certain industrial settings such as selection of a product from a range of options. Nonetheless, examples from other domains seem to show the fallacy of trying to use a single aggregated measure. For instance, healthcare professional would find a proposal to measure a person’s physical or mental health using a single score absurd, although indicators of these two health aspects may correlate.

As so-called prototypical UX measures are still lacking (at least not yet available in the ISO 9241-210, 2010 definition) and the number of empirical studies on measuring UX seems not yet large enough for meta-analysis, no claim for a summary UX score is yet made. However, it can be envisaged that it would be even more challenging to derive such a score, as the range of UX metrics seems ever growing, and so is the diversity of means to gauge them.

2.4.1. Practical implications

Categorisation of errors in usability tests remains controversial. Hornbæk and Law (2007) define a distinction between task-completion-error (errors in-task outcomes) and errors-along-the-way (e.g. slips, mistakes). In computing their correlations with task completion time, the former has a substantially lower value than the latter; this and other findings call into question the justifiability of developing a summary score. However, Sauro and Lewis (2009) argue that such a differentiation of errors may not be necessary in real-life practice. The desirable granularity of measures and unit of analysis can be very different between academic and industrial researchers; the former may tend to study a construct at its finest possible units whereas the latter may prefer more abstract entities encapsulating the details. Given the broad range of UX attributes, a taxonomy of these attributes with clear definitions grounded in theories will be very useful. The taxonomy can facilitate the selection of right attributes for the right context.

2.4.2. Research questions

Is it necessary and even possible to standardize units of measure and analysis for usability and UX? Which theoretical frameworks can well inform the development of a taxonomy that specifies unambiguously the relationships between usability and UX measures?

2.5. Implications for constructing the UX measurement models

A discussion follows of several questions in relation to developing and validating UX measurement models.

2.5.1. Why bother to have a model

A measurement model enables (or even ensures if a valid model is established and applied properly) UX measures to attain a certain level of meaningfulness and validity. Sound theoretical frameworks that define the nature and properties of a (multidimensional) UX construct (i.e. mapped onto one or more latent variables) can inform its operationalisation and manifest measures.
2.5.2. What to measure

While we do not ambitiously aim to list all possible UX measures, the issue of misspecification in measurement models (Diamantopoulos et al., 2008) is of concern. To reiterate, a measurement model specifies a relationship between a construct (latent variable) and its measures (manifest variables/indicators with data collected using, e.g. questionnaire items). The direction of the relationship can be either from the construct to the measures, known as reflective measurement, or from the measures to the construct, known as formative measurement. The former implies that all the measures in the model should be positively correlated because they are commonly caused by a construct (e.g. intelligence) whereas the latter implies that the measures in the model determine the meaning of the construct (e.g. quality of life, Bollen and Ting, 2000; e-service quality, Collier and Bienstock, 2006) and that the measures are not necessarily correlated (Diamantopoulos, 2006). These distinctions have a significant implication on scale development procedures such as item purification – a process for filtering out irrelevant items to enhance validity and reliability of a measuring scale. In constructing formative indices, discarding items showing low item-to-total correlations may actually remove those that would contribute most to the empirical meaning of the construct. However, dropping items with high inter-item correlations is an approach to eliminating multicollinearity (Howell et al., 2007). A direct consequence of misspecification is the wrong conclusion about the theoretical relationships among the constructs of interest (Diamantopoulos et al., 2008) or constructs and their indicators (e.g. the adoption of reflective indicators where formative ones would be appropriate or vice versa). Presumably, UX measurement model is formative in nature. The degree of misspecification in the literature of marketing and management is assessed to be worrisomely high (Jarvis et al., 2003; Podsakoff et al., 2006). It would be intriguing to carry out a similar assessment on the literature of UX when it reaches a reasonable size.

2.5.3. How to collect data

Before one can start to develop or test a measurement model, data need to be collected. However, the way that data are collected can influence the quality of measurement models. In particular, the presentation of psychometric scales can have a marked effect on their psychometric properties. For example, when administered on paper the same questionnaire measures may have similar or the same measurement properties as when presented on line (Harper et al., 1997; Slaughter et al., 1994) or their measurement properties may be strikingly different (Buchanan et al., 2005). Although different formats for the online presentation of items can produce similar results (van Schaik and Ling, 2003, 2007), the layout of online questionnaires can have a profound effect on measurement results in terms of ease of finding items in a survey (Norman et al., 2001) and completion time and factor structure (van Schaik and Ling, 2007).

Furthermore, when psychometric instruments are used, what is measured will obviously depend on the content of the items and its precision that are presented. More specifically, from the application of the principle of compatibility (Conner and Sparks, 2005, p. 170) to UX it follows that the more specific the items are in defining details of the experience that is measured the higher the correlation between the actual experience and the measurement is. For example, an instrument with numerous items capturing different dimensions of flow experience (an experience of total involvement) with separate items (e.g. Jackson and Marsh, 1996) about the experience of using a particular artefact would be a better measure than a flow single measuring flow in a general sense regarding the use of a larger category of artefacts or unspecified artefacts.

Where to collect data is another relevant issue: lab- or field-based. As UX is context-dependent, measures are thus sensitive to situational factors. Hence, field studies are more likely to elicit realistic experiential responses than are artificial settings in a lab.

2.5.4. When to measure

The extent of differentiation of UX constructs, which can be – but is not necessarily – related to time scale (viz. with reference to biophysiological, lower-level psychological, higher-order psychological, and social-organizational processes; Newell and Card, 1985), is an important aspect of modelling. Consider the case of aesthetics as an aspect of UX. Some research models aesthetic experience as a one-dimensional concept (e.g. ‘design aesthetic’; Cyr et al., 2006). However, other research distinguishes different components of aesthetic experience and how they are related (e.g. Lavie and Tractinsky, 2004). The implication is that if a more differentiated, but parsimonious, model can better account for UX outcomes then a more complete understanding of UX has been reached, with potential consequences for the development of theory and design guidance. In relation to time scale, some more differentiated measurements of UX (e.g. several dimensions of flow; see Jackson and Marsh, 1996) may require a smaller time scale than less differentiated measurements do (e.g. flow as a single dimension; see Choi et al., 2007) because details of a particular experience are quickly lost. Therefore, not surprisingly, more differentiated modelling may be precluded when the delay between UX and measurement increases. Consequently, the timing of experience measurement, as a basis for modelling, is related to time scale. Within-session measurements capture users’ immediate experience as it happens, whereas end-of-session and longer-term measurements capture remembered experience. Human-memory research has demonstrated that episodic memories (about particular events in time) can quickly diminish over time (Tulving, 2002). Therefore, the experience that is modelled may be very different depending on the processes of memory that are involved. Above all, all sorts of measurements – including usability and UX measures – should be rooted in sound theories. Otherwise, they are just numbers, as remarked by Kuhn (1962):

“The route from theory or law to measurement can almost never be traveled backwards. Numbers gathered without some knowledge of the regularity to be expected almost never speak for themselves. Almost certainly they remain just numbers.” (p. 44)

3. Structural models

The following issues related to structural models are discussed: the time scale, conceptual differentiation and scope, orientation with respect to time, cultural differences, and implications for selecting modelling frameworks and modelling techniques.

3.1. Time scale and scope

As mentioned earlier, the time scale of human action that is modelled can vary markedly and, consequently, can have a profound effect on the content of models of UX. Here we elaborate the four main time domains under Newell and Card’s (1985) framework: (1) natural law, where neural and biochemical processes operate, (2) psychology, where lower-level psychological processes – such as encoding into and decoding from memory – operate. (3) bounded rationality, where higher-level psychological process – such as decision-making – operate, and (4) social and organisational...
science, where social and organisational processes play a greater role. Within each time domain further levels are distinguished. Although the domain(s) under study will depend on a researcher's conceptualisation of UX, the domains most frequently addressed in UX modelling research are Domains 3 and 4. In Domain 1, physiological processes related to UX can be modelled using fuzzy logic (e.g., Mandryk and Atkins, 2007). In Domain 3, modelling may address the experience of human–computer interaction during interaction within a session. Examples include the subjective measurement of in-task subjective mental effort (e.g., van Schaik and Ling, 2009) or in-task flow experience (e.g., Pearce et al., 2005). Using a longer time frame, modelling may address UX immediately after the interaction has taken place at the end of a session. For example, subjective measures of flow may be made after users have used a web site (e.g., van Schaik and Ling, 2007). In Domain 4, users' experience of web site may be modelled in relation longer-term processes, spanning several sessions of user–computer interaction. For instance, models of technology acceptance, similar to their origins in models of rational behaviour, do not attempt to model individuals' actions on specific occasions, but instead focus on 'regularities in behaviour, consistent patterns of action, response tendencies' (Ajzen, 1988, p. 46). Examples include van der Heijden's (2003) modelling work of UX as part of a technology acceptance model. Given the time scales of human action, how are processes of human action that occur at different domains of time related? For example, in Domain 3 overall judgements of interaction can be influenced by particular episodes of the preceding experienced interaction. In particular, the most challenging experience (Pearce et al., 2005) or the most recent experience (Hassenzahl and Sandweg, 2004) may disproportionately influence overall judgements, but this is not always the case (e.g., van Schaik and Ling, 2008).

Another consideration in modelling is the scope of modelling. Firstly, confounded with time scale, the scope can be either the experience of human–computer interaction in a stricter sense – in the Domain of bounded rationality (e.g., Tractinsky et al., 2006) – or this experience in relation to broader aspects of social cognition, for example technology acceptance – in the Domain of social processes (e.g., van der Heijden, 2003). A justification for using broader scope is that even if a product is highly usable its potential benefits in terms of effective and efficient task performance will not be realised if potential users are not willing to employ it. Therefore, the modelling of acceptance can clarify how UX, together with previously established other factors, influences users' technology acceptance. Secondly, in relation to UX in a stricter sense, the scope of the experience that is modelled can be more or less comprehensive. For example, in some studies only the visual presentation of artefacts to users is involved (e.g., Tuch et al., 2009), allowing the experience of viewing and knowledge acquisition to be modelled, whereas in other studies users actually use an artefact (e.g., Hartmann et al., 2008), allowing the experience of actual use to be modelled.

3.1.1. Practical implications

Generally, different psychological and other processes operate at different time scales. Specifically, therefore, different processes of UX can be captured and modelled. More differentiated conceptualisations and measurements should produce a more complete understanding of UX. Furthermore, a narrower scope of UX modelling will produce more precise results, but may need to be considered in a broader context for their interpretation, but the opposite is true when a broader scope is chosen.

3.1.2. Research questions

What are the underlying psychological processes of UX at different time scales? How are the underlying psychological processes of UX – operating at different time scales – related?

3.2. Orientation with respect to time

UX can be modelled as a state, at a particular point or interval in time, or as a process, involving the development of this experience over time. Static models are advanced to model states of experience and are normally tested using cross-sectional data, whereas dynamic models explicitly model processes of experience and are generally tested using longitudinal data. Many studies have modelled UX as a state, examining how experience develops within an interactive session (e.g., David et al., 2007; Pearce et al., 2005). Still others have explicitly studied change in the relations between variables in experience over time in the short term (e.g., Hassenzahl, 2004; van Schaik and Ling, 2008) or longer term (e.g., Venkatesh, 2000).

3.2.1. Practical implications

Modelling UX as a state can produce an understanding of the relations between outcomes and antecedents. However, modelling UX as a process can contribute to knowledge about the development of UX as a result of time-related factors such as learning after prolonged exposure.

3.2.2. Research questions

What is the relation between state- and process models of UX? Multilevel models (Luke, 2004) can be used to investigate this relation. Furthermore, longitudinal research conducted to develop process models can be prohibitively time-consuming and costly. Because data are collected at one point or interval in time cross-sectional research may be less affected by this problem. Given these considerations, in particular, can cross-sectional data be used to develop process models or inform their development? For example, data from samples with different amounts of exposure to a particular product may reveal how UX develops and is shaped by various factors over time.

3.3. Cultural differences

Regarding cultural differences, the minds of people in different cultures have 'been programmed differently' (Hofstede, 2002; Nisbett, 2003). This is not only true in the Domain of social processes (Hofstede, 2002) – with implications for system design (Smith et al., 2004), but also in the Domains of psychology and bounded rationality (Nisbett, 2003). Therefore, irrespective of the formulation of items measuring a particular component of UX, potentially different conceptual structures may exist in different cultures. As a result people's conceptualisations of their experiences are likely to be affected and consequently the measurement structure of UX may differ between cultures. Similarly, as a consequence of cultural differences in psychological processes, the structural relations between UX components may differ in quality or strength between cultures (e.g., Teo et al., 2009).

3.3.1. Practical implications

Although more research is necessary, the assumption that measurement models and structural models of UX are equivalent may not be correct. Therefore, a psychometric instrument with good psychometric properties in the culture for which it was originally developed may not have the same properties when translated for another culture and the relations between UX components may differ.

3.3.2. Research questions

To what extent are measurement models equivalent across national cultures? To what extent are structural models of UX equivalent across national cultures?
3.4. Implications for selecting frameworks for structural modelling

The modelling framework employed in a particular study depends on the chosen time scale of events. Given a time domain, UX can be modelled using a particular orientation with respect to time (state or process). Within a particular framework, competing models can be developed and tested. An example of a framework at the level of bounded rationality is Finneran and Zhang’s (2003) Person–Artefact–Task model, originally proposed to study flow experience (e.g. used in Li and Browne (2006)), but more generally applicable to the study of UX (van Schaik and Ling, submitted for publication). There are three antecedents of UX: person (user, with stable and changeable characteristics), artefact (e.g. a web site) and task (e.g. information seeking). The interaction of these antecedents results in UX (e.g. enjoyment), which – in turn – results in consequences (e.g. positive affect). An example of a framework at the level of social processes is Davis’s (1993) and Davis and Venkatesh’s (1996) high-level description of the Technology Acceptance Model (e.g. used in van der Heijden’s modelling of UX, 2003). External variables (e.g. an artefact’s user-interface design characteristics) influence a user’s cognitive response (e.g. perceptions of ease of use), which influences the user’s affective response (e.g. attitude towards use). This response influences the user’s intention (to use the artefact), which – finally – influences his/her behaviour (e.g. rate of artefact use).

3.5. Implications for selecting structural-modelling techniques

The modelling technique(s) for a particular study depend(s) on the chosen orientation with respect to time (state or process). The modelled events depend on the chosen time scale(s) (one or more of the four domains).

3.5.1. Static modelling

The structural relations among UX variables are modelled at one point or interval in time. Research using quantitative methods builds models based on a synthesis of extant models and theories, and new insights. These are then tested using newly collected empirical data and statistical analysis techniques. Three broad categories of technique can be used to analyse most quantitative UX data: ordinary least squares (OLS; e.g. van Schaik and Ling, 2008), partial least squares (PLS; e.g. Saadé and Bahli, 2005) and (covariance-based) structural equation modelling (SEM; e.g. Cyr et al., 2006).

3.5.2. Dynamic modelling

The relations between variables are modelled over time. Statistical techniques for dynamic modelling with quantitative data include multilevel modelling (Luke, 2004), latent growth modelling (Willett and Keiley, 2000) and time series modelling (Box and Jenkins, 1976). Artificial-intelligence modelling techniques such as GOMS (Card et al., 1983) can also be used to model cyclic behaviour. SEM has been used to establish the equivalence of measurement and structural models between groups. Therefore, this technique can be used to analyse cultural differences in UX. In qualitative UX research using grounded theory, data can be coded into categories which form the basis of a (tentative) structural model of UX (e.g. Pace’s model of flow experience, 2004). More generally, explanatory displays such as matrices are recommended for presenting the relationships between antecedents, outcomes and consequences in qualitative data analysis; network displays (e.g. causal models) are recommended to show variables and their causal connections as a basis for building theory (Miles and Huberman, 1994).

3.5.3. Practical implications

OLS is more suitable for analysing smaller models with one dependent variable at a time, whereas PLS and SEM can analyse models with various dependent variables simultaneously and PLS can analyse very large models. SEM is most suitable for analysing larger samples and testing models as a whole rather than parts. Grounded theory can be particularly useful in early model development.

4. An outline of the contributions

Five papers addressing different aspects of UX are included in this special issue. Specifically, their contributions in relation to the topic of measurement and structural models of UX are presented by citing or summarizing the related descriptions in the respective papers. Some specific features of the papers are also highlighted.

Finstad addresses the usability component of a measurement model of UX. The tendency to synonymise UX and usability is especially evident in industrial practices (Ketola and Roto, 2008). However, the recognition of their distinction seems to have grown, as shown by Finstad’s (this volume) attempt to integrate usability metrics into a comprehensive UX scale under development. As a justification for the approach taken, consider Note 3 of ISO 9241-210, 2010 definition of UX, which states that “usability criteria can be used to assess aspects of UX.”

Karapanos, Zimmerman, Forlizzi and Martens “…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, the so-called experience narratives.” Time is one of the trickiest factors to address in UX research, given the evanescent nature of affective states. Karapanos et al. demonstrate the feasibility of capturing the episodic evolvement of experience with a bespoke tool.

Nacke, Grimshaw and Lindley explore the measurement of UX using physiological and subjective instruments as a function of the experimental manipulation of sound and music. Psychophysiological measures are increasingly used in evaluating UX, especially in a gameplay context where any kind of interruption should be avoided lest the flow experience can be severely disrupted. In addition to the implication of the careful use of sound and music in the game design, Nacke et al.’s findings contribute to the issue on the relationships between objective and subjective measures. Future work on identifying relevant theoretical frameworks to explicate such relationships is implied.

In O’Brien’s contribution, “The aim … was to develop and test a measurement model and a structural model of user-engagement that encompassed users’ hedonic and utilitarian motivations with respect to their impressions of an engaging online shopping encounter.” Motivation is a challenging topic in psychology; it is hard to define or measure. O’Brien assumes the challenge to explore how the two distinct types of motivation are related to a cluster of related variables. Outcomes of this and similar efforts can contribute to the construction of a taxonomy of UX attributes.

Hassenzahl, Diefenbach and Göritz “…offer a structural model of positive experience, which differentiates experience based on the psychological needs fulfilled through technology use. In addition, we gain first insights into the process, which links experience to product perception, and we provide an alternative approach to the measurement of experience …” As UX is multifarious, scoping as a critical challenge involves deciding on which variables to be included in a model. Hassenzahl et al.’s study can well illustrate how a relatively complicated UX model is constructed and validated with appropriate measures and sophisticated statistical techniques.

5. Conclusion

The topic of modelling UX can divide the UX community into two major camps. One camp – ‘modelling-friendly’ – is represented
by those who are strongly convinced that it is necessary, plausible and feasible to measure UX at its finest possible units. Another camp – ‘modelling-sceptical’ – is represented by those who express their doubts about the necessity and utility of measuring UX attributes. Apparently, there are some who affiliate with both camps. Measurability of constructual concepts is indeed a longstanding argument in social science (Bulmer, 2001). We espouse Lord Kelvin’s dictum “to measure is to know”. Nonetheless, with our awareness of the risk of extreme reductionism, we emphatically point out two important criteria – meaningfulness (validity) and reliability – of UX measures. Specifically, meaningfulness should be grounded in a deep understanding of the related theoretical frameworks on experience, consciousness, memory, affect, emotion, aesthetics, social collaboration and interaction, to name just a few. Reliability is hinged upon the consistent use of the tool(s) and protocol(s) used for data collection. Obviously, measuring is important, but not enough. This is because even if one knows how to measure (for example) current and voltage, one still needs to know the laws of electricity in order to design a circuit. Therefore, even if UX measurement is successful, it becomes more useful when structural models are developed that help building the theoretical understanding of (causal) relations between UX constructs and design characteristics as a basis for informing (practical) system design. Furthermore, reflecting upon the review process of this special issue, we echo the concern of Cairns (2007) about a widespread lack of understanding of (causal) relations between UX constructs and design characteristics as a basis for informing (practical) system design. Conversely, a danger would be that lower importance, priority and attention may be given to statistical methods. While we do not intend to make any claim that one type of method is superior to the other, we intend to draw the attention of HCI educators to this concern.

Acknowledgement

We thank all the reviewers who made this Special Issue possible:

<table>
<thead>
<tr>
<th>Javier Andres</th>
<th>Kasper Hornbaek</th>
<th>Don Shin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bargas-Avlá</td>
<td>Ebba Thora</td>
<td>Stephen Smith</td>
</tr>
<tr>
<td>Martin Beer</td>
<td>Hvannberg</td>
<td>Mark Springett</td>
</tr>
<tr>
<td>Regina Bernhaupt</td>
<td>Nette Sivari</td>
<td>Alistair Sutcliffe</td>
</tr>
<tr>
<td>Nigel Bevan</td>
<td>Vassilis Kostakos</td>
<td>Sharon Y. Tettegah</td>
</tr>
<tr>
<td>Samit Bhattacharya</td>
<td>Sari Kujala</td>
<td>Manfred Thuering</td>
</tr>
<tr>
<td>Béatrice Cahun</td>
<td>Effie L.-C. Law</td>
<td>Alexandre N. Tuch</td>
</tr>
<tr>
<td>Georgios Christou</td>
<td>Gittie Lindgaard</td>
<td>Phil Turner</td>
</tr>
<tr>
<td>Oscar De Bruijn</td>
<td>Zhengjie Liu</td>
<td>Kaisa Vaananen-Vainio-Mattila</td>
</tr>
<tr>
<td>James Patrick</td>
<td>Rod McCull</td>
<td>Paul van Schaik</td>
</tr>
<tr>
<td>Downey</td>
<td>Heather Lynn</td>
<td>Lex van Velsen</td>
</tr>
<tr>
<td>Asbjørn Flåstad</td>
<td>O'Brien</td>
<td>Martina Ziele</td>
</tr>
<tr>
<td>Qin Gao</td>
<td>Talla Porat</td>
<td>Jürgen Ziegler</td>
</tr>
<tr>
<td>Joely Gardner</td>
<td>David Pythrech</td>
<td></td>
</tr>
<tr>
<td>David Geerts</td>
<td>Pei-Luen Patrick Rau</td>
<td></td>
</tr>
<tr>
<td>Dong-Han Ham</td>
<td>Virpi Roto</td>
<td></td>
</tr>
<tr>
<td>Mari lien Herselman</td>
<td>Martin Schrepp</td>
<td></td>
</tr>
</tbody>
</table>

References

Draper, S.W., 1993. The notion of task in HCI. In: Proceedings of CHI’93, Amsterdam, the Netherlands, pp. 207–208.

Effie L.-C. Law
Department of Computer Science, University of Leicester, LE1 7RH, Leicester, UK
E-mail address: elaw@mcs.le.ac.uk

Paul van Schaik
Psychology Subject Group, School of Social Sciences and Law, Teesside University, UK
E-mail address: p.van-schaik@tees.ac.uk