

Chapter 1

Introduction

48% of returned products are not attributed to a violation of product specifications (Den Ouden et al., 2006). This finding was the initial motivation for this research project. Brombacher, den Ouden and colleagues (e.g., Den Ouden et al., 2006; Brombacher et al., 2005; Koca et al., 2009) found that an alarmingly increasing number of returned products, in 2002 covering 48% of returned products, are technically fully functional, i.e. according to specifications, but they are returned on the basis of failing to satisfy users' true needs (28%), or purely because of users' remorse (20%) (Den Ouden et al., 2006). Brombacher et al. (2005) introduced the term 'Soft Reliability' to refer to these situations where "*in spite of meeting with the explicit product specifications, a customer explicitly complains on the (lack of) functionality of the product*".

How is this finding different from Suchman's well-known case at XEROX in the '80s where users were found to have problems in using a feature-rich photocopier (c.f., Suchman, 2006)? While product designers were aiming at improving the instructions for using the product, Suchman argued that learning is an inherently problematic activity, and suggested that "*no matter how improved the machine interface or instruction set might be, this would never eliminate the need for active sense-making on the part of prospective users*" (Suchman, 2006, , p. 9). Since the '80s, a wealth of theories, methods, and design guidelines have been developed in the field of Human-Computer Interaction with the aim of making products more easy to learn and use in the long run. Thus, one might wonder, do Brombacher's and den Ouden's findings replicate what was found at XEROX almost 30 years ago, or do they introduce a new and as yet unaddressed problem? Should these consumer complaints be attributed to bad design practices, to apparently inescapable interaction flaws in first-time use, or do they suggest a new emerging problem in the user acceptance of interactive products?

Den Ouden et al. (2006) identified a number of trends in the consumer electronics (CE) industry that have resulted in radical differences of the current market in comparison to that in the '90s. They argued that the emphasis in the CE industry has shifted from the production of high volumes at competitive prices to the introduction of highly innovative products at higher prices. This leads to a shift in the

main uncertainty in new product development projects; while in the '90s the uncertainty related to the technology in relation to cost-effective mass production, in the current market the dominant uncertainty relates to the attractiveness of the product and users' expectations about the product functions.

These trends are reflected in the development of the field of Human-Computer Interaction, from the study of usability as a critical factor to the acceptance of interactive products, to a more holistic understanding of users' experiences with interactive products, leading to the study of new concepts like pleasure (Jordan, 2000), fun (Blythe et al., 2003), aesthetics (Tractinsky et al., 2000) and hedonic qualities (Hassenzahl, 2004). While a wealth of techniques and methods exist for ensuring the usability of interactive products, research on user experience evaluation methods is only at its infancy. This book aims at highlighting methodological issues in user experience evaluation and proposes a number of methods for inquiring into users' experiences with interactive products.

1.1 From Usability to Experience

The field of Human-Computer Interaction was for a long time identified as the field of usability engineering. Usability was seen as critical to user acceptance and a wealth of principles (e.g., Norman, 1988), design guidelines (e.g., Nielsen and Bellcore, 1992) and evaluation techniques (e.g., Dix et al., 2004) have become instrumental in the development of usable products. The field of usability engineering readily acknowledged the dual nature of the usability concept: its *objective* and *subjective* side. One of the most dominant definitions of usability, for instance, the ISO 9241-11 standard (1996) defines usability as

“the extent to which a product can be used by specific users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”.

Effectiveness represents the accuracy and completeness with which users achieve certain goals and is typically measured through observed error rates, binary task completion, and quality of the outcome for a given task (see Frøkjær et al., 2000; Hornbæk and Law, 2007). Efficiency can be characterized as effectiveness in relation to resources spent and is typically measured through task completion time and learning time. One can note that both these components, effectiveness and efficiency represent the objective side of usability. The third component, user satisfaction, represents users' comfort in using and overall attitude to the product and is typically measured through psychometric scales on overall preference, product quality perceptions and specific attitudes towards the interface (see Hornbæk and Law, 2007).

An assumption underlying the distinction between the subjective and objective side of usability was that these two would strongly correlate. Nielsen and Levy (1994), in a meta-analysis of a selected set of 57 studies found that in 75% of the cases, users' overall preference was strongly related to overall performance. These findings have, however, been repeatedly questioned in subsequent studies suggesting

that subjective perceptions of usability are generally not correlated with objective measures and seem to measure something else than merely effectiveness and efficiency (Frøkjær et al., 2000; Hornbæk and Law, 2007; Kissel, 1995; Bailey, 1993). This limited view on user satisfaction as a consequence of objective performance was criticized by Hassenzahl et al. (2000):

“...it seems as if satisfaction is conceived as a consequence of user experienced effectiveness and efficiency rather than a design goal in itself. This implies that assuring efficiency and effectiveness alone guarantees user satisfaction.”

Subsequently, a set of studies tried to explain these observed discrepancies between objective and subjective usability. Kurosu and Kashimura (1995) asked participants to rate several Automatic Teller Machine (ATM) designs on both functional and aesthetic aspects. They found apparent usability, i.e. subjective judgments on usability, to correlate more strongly with aesthetic judgments than with the systems' inherent usability, i.e. objective design parameters that were expected to affect users' performance in using the systems. Tractinsky (1997) and Tractinsky et al. (2000) replicated this study and found that these effects persisted both across different cultural backgrounds, as well as after participants had experienced the systems.

These early findings suggested that users' experiences with products go beyond the effectiveness and efficiency in product usage. Consequently, the field of Human-Computer Interaction quested for new *concepts, measures* and *methods* in capturing a more holistic view on user experience. This development has gone hand-in-hand with a shift in the contexts of study, from *professional* to *personal* (e.g., Jordan, 2000) and *social* (e.g., Forlizzi, 2007; Markopoulos et al., 2004), and in the design paradigm from *product* to *experience design* (e.g., Buxton, 2007; Zimmerman et al., 2007; Forlizzi et al., 2008).

1.2 Two Distinct Approaches in User Experience Research

User experience has become central to the design and evaluation of interactive products. It reflects a paradigm shift in the subject of product design and evaluation. Buxton (2007) argues the following:

“Ultimately, we are deluding ourselves if we think that the products that we design are the “things” that we sell, rather than the individual, social and cultural experience that they engender, and the value and impact that they have. Design that ignores this is not worthy of the name”

However, user experience research is often criticized for at least two things: a) for the lack of a commonly agreed definition of the notion of experience, and b) for being identical, conceptually or methodologically, to traditional usability research. Indeed, Hassenzahl (2008) and Wright and Blythe (2007), some of the strong proponents of user experience research, criticize the use of the term *user experience* in cases where the focus still lies in traditional usability evaluation, thus

reducing the richness of experience to behavioral logs and task-focused evaluations (Wright and Blythe, 2007). As Hassenzahl (2008) argues:

“While UX seems ubiquitous in industry, a closer look reveals that it is treated mainly as a synonym of usability and user-centered design”

Conversely, as Hassenzahl (2008) argues, academics “emphasize the differences between traditional usability and user experience”. A number of frameworks have tried to conceptualize how experiences are formed (e.g., Forlizzi and Ford, 2000; Wright and McCarthy, 2004; Norman, 2004; Hassenzahl, 2008) and tentative definitions of experience have been proposed (Forlizzi and Battarbee, 2004; Hassenzahl and Tractinsky, 2006; Hassenzahl, 2008).

Hassenzahl (2008) defines user experience as “*a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service*” which is “a consequence of users’ internal state (e.g. predispositions, expectations, needs, motivation, mood), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality), and the context within which the interaction occurs (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use)” (Hassenzahl and Tractinsky, 2006). One may argue that such a definition, while being perhaps the best of what the field of user experience can offer at the moment, is far from being mature or useful for grounding measures, methods and principles in the design and evaluation of interactive products. Yet, a common ground has been established among various disciplines and schools of thoughts in the emerging field of user experience, perhaps due to “a history of use of the term in ordinary conversation and philosophy discourse” (Wright and Blythe, 2007). A number of researchers have tried to identify the dominant schools of thought and several classifications have been proposed (e.g. Battarbee and Koskinen, 2005; Hassenzahl and Tractinsky, 2006; Blythe et al., 2007).

We employ the distinction from Blythe et al. (2007) between *reductionist* approaches that have their roots in cognitive psychology, and *holistic* approaches that are grounded in pragmatist philosophy and phenomenology. As it will become apparent in section 1.4, we are primarily interested in distinct issues that these two approaches pose when one is concerned about methodology for understanding user experiences.

1.2.1 Reductionist Approaches

Reductionist approaches in user experience maintain a similar paradigm to *usability* (ISO, 1996) and *Technology Acceptance* research (see Venkatesh et al., 2003) in trying to identify distinct psychological constructs and propose and empirically test causal relations between them.

One of the first and well cited studies in user experience, grounded on reductionism, is that of Tractinsky (1997). Tractinsky was puzzled by the findings of Kurosu and Kashimura (1995) who suggested that subjective perceptions of usability relate

more to the beauty of the product than to its actual, i.e. inherent, usability. This finding highlighted the importance of aesthetics in interactive products, an opinion that found Tractinsky resonant with. Tractinsky, however, predicted that this might not pertain over different cultures, taking into account that the study of Kurosu and Kashimura (1995) was conducted within the Japanese culture which is known for its aesthetic tradition. Tractinsky (1997) replicated the experiment using the same stimuli, but now in an Israeli context. His initial prediction was not confirmed as the findings of Kurosu and Kashimura (1995) were reconfirmed in this alternative setting. One possible criticism of both studies could be that the user judgments were elicited merely on the basis of the visual appearance of the interface without experiencing the systems. In a subsequent study, Tractinsky et al. (2000) elicited users' perceptions both before and after interacting with a computer simulation of an interface of an Automatic Teller Machine (ATM). The results suggested that the aesthetics of the interface also impacted the post-use perceptions of usability. Subsequent work has supported the dominance of beauty in users' preferences (e.g., Schenkman and Jönsson, 2000; Lindgaard and Dudek, 2003; Lindgaard et al., 2006; Tractinsky et al., 2006; Hekkert and Leder, 2008) and have provided further insight into users' inferences between aesthetics and usability (e.g., Hartmann et al., 2008).

Hassenzahl (2004) wanted to further inquire into the nature of beauty in interactive products. He developed a theoretical model (Hassenzahl, 2005) that distinguishes between objective parameters, product quality perceptions and overall evaluations. Based on this model he understood beauty as "a high-level evaluative construct comparable to (but not identical with) other evaluative constructs, such as goodness or pleasantness" (Hassenzahl, 2004, , p. 323) and perceived usability as a bundle of lower level judgments reflecting product quality perceptions. He distinguished between two quality perceptions: *pragmatic* and *hedonic*. Pragmatic quality, he argued, 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*, i.e. the product's ability to stimulate and enable personal growth, and *identification*, i.e. the product's ability to address the need of expressing one's self through objects one owns. He further distinguished between two overall evaluative judgments: *goodness* and *beauty*. Contrary to (Tractinsky et al., 2000), he found minimal correlation between pragmatic quality, i.e. usability, and beauty. Beauty was found to be a rather social aspect, largely affected by identification; pragmatic quality, on the contrary, related to the overall judgment of goodness. In a similar vein, Tractinsky and Zmiri (2006) distinguished between satisfying and pleasant experience. They found perceptions of usability to be better predictors of satisfying rather than pleasant experience while perceptions of products' aesthetics to be better predictors of pleasant rather than satisfying experience. Hassenzahl's (2004) model of user experience has also been further supported by subsequent research (e.g., Mahlke, 2006; Schrepp et al., 2006; Van Schaik and Ling, 2008).

Mahlke and Thüning (2007) provided a comprehensive framework linking product quality perceptions to emotional reactions and overall evaluative judgments. Their findings supported Hassenzahl's 2004 distinction between goodness and

beauty, with goodness relating primarily to instrumental qualities, e.g. usefulness and ease-of-use, and beauty relating primarily to non-instrumental qualities such as the visual aesthetics and haptic quality (Mahlke, 2006). Desmet (2002), grounded on Russell's (1980) model of affect, developed a tool for measuring emotional responses to products and established a framework that relates aesthetic response to meaning (see Desmet and Hekkert, 2007). Fenko et al. (2009) studied how the dominance of different sensory modalities such as vision, audition, touch, smell and taste develops over different phases in the adoption of the product such as when choosing the product in the shop, during the first week, after the first month, and after the first year of usage.

1.2.2 *Holistic Approaches*

Holistic approaches are rooted in pragmatist philosophy and phenomenology. They criticize reductionist approaches in that they reduce the complexity and richness of user experience to “a set of manipulable and measurable variables” and impose “abstract models and classifications onto rich and complex models like affect and emotion” (Wright and Blythe, 2007) (see also Hassenzahl, 2008). Similarly, Suri (2002) argues that “measurement, by its nature, forces us to ignore all but a few selected variables. Hence, measurement is useful when we are confident about which variables are relevant”. She argues that designers are concerned about developing new products and for new contexts and thus no such understanding exists about how design attributes and contextual details interact in given contexts, and proposes alternative methods, such as that of *experience narratives*, for inquiring into how product meaning and value emerges in given contexts.

From a theoretical point of view, holistic approaches have contributed a number of frameworks describing how experience is formed, adapted, and communicated in social contexts.

Forlizzi and Ford (2000) provided a framework that attempts to describe how experience transcends from unconsciousness to a cognitive state and finally becomes “an experience”, something memorable that can also be communicated in social interactions. They identified four modes or dimensions of experiencing: sub-consciousness, cognition, narrative and storytelling. Sub-consciousness represents fluent experiences that do not compete for our attention. Cognition represents experiences that require our attention, e.g. learning to use an unfamiliar product. Narrative represents “experiences that have been formalized in the users' head: ones that force us to think about and formulate what we are doing and experiencing”. Forlizzi and Ford (2000) suggest that a product's set of features and affordances offers a narrative of use. Storytelling, represents the subjective side of experience: “a person relays the salient parts of an experience to another, making the experience a personal story”. Forlizzi and Ford (2000) argue that through these particular sense making users attach meaning and personal relevance to situation, “creating life stories and stories of product use”. Forlizzi and Ford (2000) subsequently identify ways of shifting across these four modes of experiencing. One might migrate

from a cognitive to a sub-conscious experience, for instance by learning how to use a product. Reversely, a fluent experience may shift to a cognitive if a user encounters something unexpected in her interaction with the product and is forced to think about it. A narrative experience can shift to a cognitive one when one “is forced to challenge his own thinking that has been solidified in her perceptions, attitudes, and beliefs”. An experience might also shift from a sub-conscious state to story-telling as she “schematizes it, communicates it and add levels of meaning”.

Forlizzi and Battarbee (2004) modified this framework to include the concept of co-experience proposed by Battarbee (2003). Battarbee and Koskinen (2005), further elaborated on the social mechanisms that lift or downgrade experiences as they participate in people’s social interactions. They identified three distinct mechanisms: *lifting up experiences*, *reciprocating experiences*, *rejecting and ignoring experiences*. First, people may “lift things from the stream of events”, considering them as meaningful enough to be communicated in social settings. Secondly, recipients of communicated experiences may acknowledge the described experience as personally relevant and respond to it by telling their own, similar experiences. Finally, experiences communicated in social settings may be rejected or downgraded by others, eventually altering the dominance of the given experience for the person who chose to communicate it.

McCarthy and Wright (2004) distinguished between four threads of experience: *compositional*, *sensual*, *emotional*, and *spatio-temporal*. The compositional thread concerns the way that different elements of experience form a coherent whole. It refers to “the narrative structure, action possibility, plausibility, consequences and explanations of actions”. The sensual thread relates to “the concrete, palpable, and visceral character of experience that is grasped pre-reflectively in the immediate sense of a situation”. The emotional thread refers to value judgments (e.g., frustration and satisfaction) that ascribe importance to other people and things with respect to our needs and desires”. Lastly, the spatio-temporal thread “draws attention to the quality and sense of space-time that pervades experience”. McCarthy and Wright (2004) pinpoint that while these are positioned as distinct components of experience they should be seen as intrinsically connected with each other.

Next to the four threads of experience, McCarthy and Wright (2004) described how sense-making takes place in the development of experience by decomposing it into six processes: *anticipating*, *connecting*, *interpreting*, *reflecting*, *appropriating*, and *recounting*. Anticipation refers to users’ expectations and imagined possibilities that are grounded in prior experience. In connecting, users make an instant judgments referring to the immediate, pre-conceptual and pre-linguistic sense of a situation. In interpreting, users work out what’s going on and how they feel about it. In reflecting users examine and evaluate what is happening in an interaction and the feelings of frustration or pleasure that are part of the experience. In appropriating, users evaluate how the new experience relates to prior experiences, and in recounting, users communicate the experienced situation to others and reinterpret the experience as it participates in storytelling.

1.3 Diversity in User Experience

While one may quickly note that holistic approaches emphasize the uniqueness of experience across different situations and people, both approaches to user experience have readily accepted that diversity in users' experiences is prevalent. A wealth of empirical studies, grounded on reductionism, acknowledge and tackle diversity empirically (e.g. Hassenzahl and Ullrich, 2007; Mahlke and Lindgaard, 2007).

This section introduces the notion of diversity in user experience. We introduce a framework of diversity in subjective judgments and identify four different sources of diversity in users' experiences with interactive products. Only later, in section 1.4 we will introduce the methodological differences between the two approaches, reductionist and holistic, in accounting for diversity.

1.3.1 A Framework of Diversity in Subjective Judgments

Diversity was readily accepted in the HCI field as a key issue. Not all users like the same things and different product qualities suffice in different situations (e.g. Cooper, 1999). But, other fields have been constructed on the assumption of homogeneity across different individuals. In the field of psychophysics, for example, the principle of *homogeneity of perception* states that different participants will more or less agree on perceptual judgments such as how much noise, or blur, an image contains, or how much friction, or inertia, one may find in a haptic control. This assumption has been instrumental in the development of respective statistical techniques; for instance, Multi-Dimensional Scaling (Green et al., 1989; Martens, 2003), motivated by this principle, assumes that judgments of different individuals may be visualized in a shared K-dimensional configuration of stimuli, for which the coordinates of the stimuli in the configuration space along different axes can be monotonically related to the observed attribute ratings of the participants.

It seems natural to accept that while different individuals might agree on low-level perceptual judgments, e.g. friction in a haptic control, a relative disagreement would be found as one moves to more cognitive judgments, e.g. the degree to which this haptic control is perceived as playful. Indeed, in an exploratory study we tried to inquire into whether people agree on *product character* judgments (Janlert and Stolterman, 1997) of a physical rotary knob. A project done by Bart Friederichs in his Master's studies, in collaboration with BMW, aimed to prove that physical controls, i.e. rotary knobs, can be designed so that they comply with the personality of a given car. But, how much do different users agree on high level product character judgments of haptic controls? Eighteen participants experienced fifteen different haptic scenes programmed in a haptic feedback knob and rated them in terms of four different bipolar scales: *Dynamic - Static*, *Agile - Plump*, *Playful - Serious*, and *Luxurious - Austere*. One may quickly note that the last two judgments, those of playfulness and luxury, are higher level judgments, allowing for more disagreement on what one considers playful or luxurious. We applied a uni-dimensionality test, by applying two Multi-Dimensional Scaling solutions on the four ratings: a 1D and a 2D. A χ^2 test, showed that the first two judgments, which we assumed as lower

level, perceptual judgments, were indeed more uni-dimensional, i.e. they displayed higher consistency across different individuals.

In the field of design, Csikszentmihalyi and Rochberg-Halton (1981) asked individuals to select personally relevant objects located in their homes and describe what makes them special. Csikszentmihalyi and Rochberg-Halton (1981) found that the value of such objects did not lie in some objectively defined quality, e.g. uniformly appreciated aesthetics, but to the personal meaning that people attached to these objects and how these participated in their social lives and creation of self-identity. These results suggest that while we may all agree in perceptual judgments, e.g. color of a given product, these judgment have far less power in predicting our preferences in comparison to higher level judgments, e.g. its beauty. Hofstede (2001) suggested that human perceptions may exist at three different levels: a) some might be uniquely personal, having significance for an individual because of his or her own associations; b) others might have significance to a specific social or cultural group through shared meaning; and c) others are universal, related to human nature at an innate or fundamental level.

In Karapanos et al. (2008b) we argued that diversity may exist at two different stages in the formation of an overall evaluative judgment (see figure 1.1). Perceptual diversity lies in the process of forming product quality perceptions (e.g. novel, easy to use) on the basis of product features. For instance, different individuals may infer different levels on a given quality of the same product, e.g. disagree on its novelty. Evaluative diversity lies in the process of forming overall evaluations of the product (e.g. good-bad) on the basis of product quality perceptions. For instance, different individuals may form different evaluative judgments even while having no disagreement on the perceived quality of the product, e.g. both might think of it as a novel and hard-to-use product, but they disagree on the relative importance of each quality.

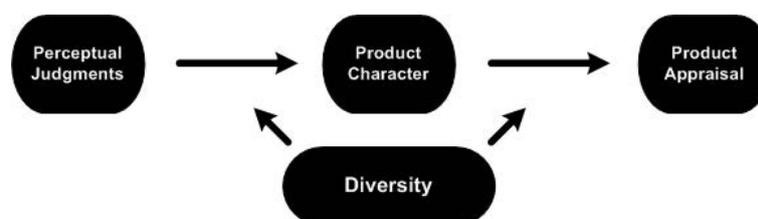


Fig. 1.1 A modification of Hassenzahl's 2005 framework, highlighting diversity at two stages in forming overall judgments about products.

Considering the second stage, i.e. evaluative diversity, one might assume a certain universal hierarchical structure on the relative importance of different qualities. For instance, Jordan (2000) drew on Maslow's (1946) theory of human needs to propose a fixed hierarchical structure of the relative importance of functionality, usability and pleasure in the context of Human-Computer Interaction. According to this, a product has to provide useful and usable functionality before hedonic aspects,

such as beauty and stimulation, can take effect. Similarly to the field of psychology where Maslow's theory was readily adopted while lacking empirical evidence, Jordan's fixed hierarchy has become widely popular in the field of user experience, but, to our knowledge, no empirical studies have attempted to confirm or disprove the framework. Contrary to Jordan (2000), Hassenzahl (2006) assumes the importance of these different qualities to vary with several contextual factors.

1.3.2 Four Sources of Diversity in User Experience

In the field of user experience, a wealth of empirical studies have tried to tackle how the relative importance of different product qualities on users' overall evaluations and preferences, is modulated by a number of contextual aspects. In figure 1.2 we highlight four different sources of diversity that user experience research has been concerned with.

First, individual differences (e.g. human values Schwartz, 1992) moderates the importance individuals attach to different qualities of an interactive product (e.g. Desmet et al., 2004; Karapanos and Martens, 2007); while some might prefer playful and stimulating products, others might value simplicity and austerity. Second, the type of the product matters (e.g. Jordan and Persson, 2007); while a playful interaction might be crucial for the success of a computer game, the same quality might be perceived as inadequate for professional software. Third, even for the same product, the way individuals use it differs across situations and this impacts the importance that they attach to different qualities (e.g. Mahlke and Lindgaard, 2007; Hassenzahl and Ullrich, 2007; Hassenzahl et al., 2008); the same mobile phone could be used for exploring the available ring tones or to make an emergency call.

A fourth aspect, which is mostly overlooked so far, is the systematic change of experience over time. As individuals use a product, their perception of the qualities of the product will change (e.g. von Wilamowitz Moellendorff et al., 2006; Karapanos et al., 2008a, 2009c; Fenko et al., 2009). For example, they get used to it, which eventually changes their perception of its usability; at the same time it excites them much less than initially. Even more interestingly, at different phases of use they will evidently attach different weights to different qualities. In their first interactions with a product they may focus on its usability and stimulation. After they use it for some time, they might become less concerned about its usability, and other aspects of the product such as novel functionality or communication of a desired identity towards others become more important.

All these factors, the individual, the product, the situation, and time modify the importance of the qualities for a satisfying experience with an interactive product. In this view, Jordan's (2000) hierarchy of consumer needs could be seen as a particular, context-dependent prioritization of needs (Hassenzahl, 2006).

This section introduced the notion of diversity, which mostly concerns in the user experience field the relative dominance of different product qualities while forming overall judgments about interactive products. The next section introduces the methodological debate in accounting for diversity between the two dominant

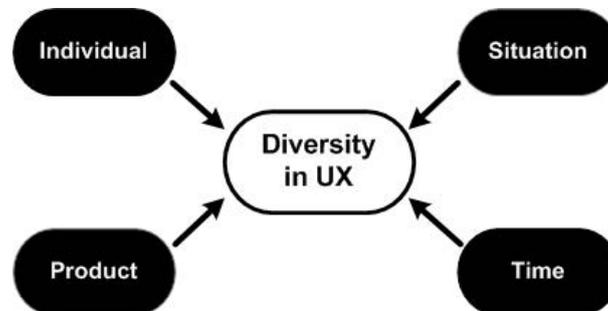


Fig. 1.2 Four different sources of diversity in user experience. These are assumed to modulate the dominance of different product qualities on users' experience and overall evaluative judgments.

approaches in user experience research, it discusses the limitations of each approach, and argues for a hybrid methodological paradigm that shares values from both approaches, the holistic and the reductionist.

1.4 Methodological Issues in Accounting for Diversity

Reductionist approaches typically employ psychometric scales (e.g. Likert, 1932; Osgood et al., 1957) in measuring latent psychological constructs that the researchers consider as relevant for a certain context. Such approaches are grounded on the assumption that people may summarize experiences in overall evaluative judgments (Hassenzahl, 2008). When a product is associated with a certain experience, the value of the experience will be partially attributed to the product.

Proponents of the holistic paradigm criticize reductionist approaches for not capturing "grey areas and moment-by-moment decision making" (Wright and Blythe, 2007). Instead, they propose holistic and situated techniques that are motivated by the need to establish an empathic relation with the user (see Wright and McCarthy, 2008). In one of the most popular techniques, the *cultural probes* (Gaver et al., 1999), participants are typically provided with a camera and a set of abstract objects to inspire them in capturing their experiences, dreams and aspirations related to a given topic of interest. Suri (2002) proposed the elicitation of short essays, i.e. *experience narratives*, in understanding the frustration and enjoyment that is experienced while interacting with products. Blythe et al. (2002) proposed a technique called *technology biographies* in which participants are asked to reminisce how their relationship with technology has changed through their lives. In the creation of *Anticipation and reflection interviews*, (Blythe et al., 2006) acknowledge that experience is not limited to a given interaction instance but extend to the process of forming expectations of and reflecting on experiential episodes.

One might argue that these two approaches have distinct goals, being *evaluative* or *inspirational*. In this sense, holistic approaches to evaluation serve to inspire design solutions (Hassenzahl, 2008); evaluation is here understood as idea generation

(Hornbæk, 2008). As Hassenzahl (2008) puts it “[designers] are able to build ideas from anecdotal observations and loose associations”. In contrast, reductionist approaches to evaluation serve to assess the value of a design, to compare multiple designs, or to develop theory and criteria to support evaluation.

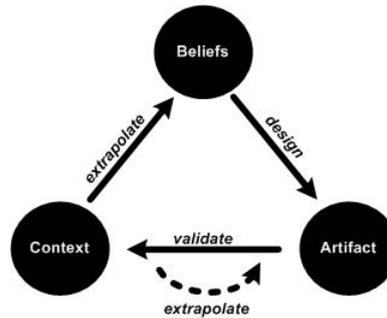


Fig. 1.3 A simplified framework of the design process, derived from (Martens, 2009b, personal communication). According to this framework, the designer forms beliefs about the potential consequences of use of a given product in a given context, grounded on empirical knowledge. These beliefs are in turn externalized to artifacts and the artifacts are validated on a number of criteria that have been derived from prior beliefs, or empirical insights from the previous design iteration.

We argue against this distinction between evaluative and inspirational goals in product evaluation. It is argued that current evaluation practices that embrace this distinction lead to inefficacy in the evaluation of interactive products. This is reflected in figure 1.3 which depicts a simplification of a framework of the design process initially proposed by (Martens, 2009b, personal communication). The framework identifies three distinct activities in an iterative design process: a) extrapolating knowledge from context and forming beliefs about the potential consequences of use of a given product within a given context, b) transforming (implicit) beliefs to explicit knowledge through the design of artifacts, and c) validating the designed artifacts in context.

According to this view, while holistic approaches serve to extrapolate knowledge in forming or adapting the designer’s beliefs, reductionist approaches serve to validate the designed artifacts using a number of criteria that have been derived from prior beliefs, or empirical insights from the previous design iteration. This leads to a number of limitations in the evaluation process. First, reductionist approaches to evaluation through pre-defined measurement scales may miss potentially relevant concepts due to a failure of researchers to recognize their relevance, but may also result in less meaningful information when participants cannot interpret the questions with respect to their own context. Second, holistic approaches such as the experience narratives, provide rich information in all aspects that surround an experience but may create a risk of focusing on idiosyncratic experiences while failing

to estimate its dominance and probability of occurrence over the total population of users.

In combination, rich situated insights derived from holistic methods may inform the validation of artifacts while reductionist approaches to evaluation may quantify the importance of a given experience and thus minimize the risk of overemphasize interesting but rare experiences.

We argue for a missing link between *validation* and *extrapolation* (see figure 1.3). Below, we describe how this is addressed in relation to the two research foci of this manuscript: understanding interpersonal diversity in users' responses to conceptual designs, and understanding the dynamics of experience over time.

1.4.1 Understanding Interpersonal Diversity through Personal Attribute Judgments

Den Ouden (2006) revealed that the majority of soft reliability problems related to the concept design phase and were particularly rooted in design decisions relating to the product definition. This insight suggests that design decisions made early in the design process may not be adequately grounded on empirical user insights.

Traditional approaches to measuring users' responses to artifacts lie in the a-priori definition of the measures by the researchers. This approach is limited in at least two ways when one is concerned with capturing the richness of and diversity in user experience. First, the a-priori definition of relevant dimensions is inherently limited as researchers might fail to consider a given dimension as relevant, or they might simply lack validated measurement scales, especially in developing fields such as that of user experience where radically new constructs are still being introduced. Secondly, one could even wonder whether rating a product on quality dimensions that are imposed by the researcher is always a meaningful activity for the user, for example when the user does not consider a quality dimension as relevant for the specific product. There is increasing evidence that users are often unable to attach personal relevance to the statement provided in psychometric scales due to a failure to recall experiential information that relates to the statement or due to lengthy and repetitive questioning. Larsen et al. (2008b) reviewed a number of studies employing psychometric scales in the field of Information Systems. They found for the majority of studies the semantic similarity between items to be a significant predictor of participants' ratings ($.00 < R^2 < .63$). In such cases, they argued participants are more likely to have employed *shallow processing* (Sanford et al., 2006), i.e. responding to surface features of the language rather than attaching personal relevance to the question.

An alternative approach to predefined questionnaires lies in a combination of structured interviewing, that aims at eliciting attributes that are personally meaningful for each participant, with a subsequent rating process performed on the attributes that were elicited during the interview. This approach aims at increasing the diversity and relevance to the individual of concepts that are measured, thus resulting in richer insights. However, the techniques required for the quantitative analysis of

such information become significantly more complex. We propose two techniques for the analysis of personal attribute judgments.

1.4.2 Understanding the Dynamics of Experience through Experience Narratives

Product evaluation practices have traditionally been focusing on early interactions. As a result, we have been mostly concerned about product qualities relating to the initial use. Den Ouden et al. (2006), however, highlighted that the reasons for product returns span a wider range of aspects than just problems related to their learnability and usability. Moreover, a number of recent trends are highlighting the importance of longitudinal evaluation practices. First, legislation and competition within the consumer electronics industry has resulted in an increase in the time-span of product warranties. This has resulted in an alarmingly increasing number of products being returned on the basis of failing to satisfy users' true needs (Den Ouden et al., 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 (Karapanos et al., 2009c). Thus, the overall acceptance of a product shifts from the initial purchase to establishing prolonged use.

Traditional approaches in the study of dynamics of experience over time typically employ validated measurement and structural models across different phases in the adoption of a system (e.g. Venkatesh and Davis, 2000; Venkatesh and Johnson, 2002; Kim and Malhotra, 2005). For instance, Venkatesh and Davis (2000) employed the Technology Acceptance Model (Davis et al., 1989) at three instances in the adoption of an information system at a workplace: before the introduction of the system (inquiring into users' expectations), right after the introduction of the system, and three months after the introduction. These approaches, while being widely validated and well-cited in the field of Information Systems, are hindered by a number of limitations, at least in developing fields such as that of user experience.

An assumption inherent in this approach is that the relevant latent constructs remain constant, but their perceived value and relative dominance might change over time. But, especially in developing fields such as user experience, substantial variations might occur over time even in what constructs are relevant to measure. Some constructs, e.g. novelty, might cease to be relevant while others, such as supporting daily rituals, enabling personalization, and communicating a positive image about one's self (see Karapanos et al., 2009c), 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 is often observed in studies of user acceptance and user experience where the a-priori defined constructs account for a limited amount of the variance in the predicted variable, being it preference judgments, dissimilarity judgments or attitude towards behavior (see Venkatesh et al., 2003). Secondly, it may also lead to distorted data as individuals might fail to interpret the personal relevance of a given scale item to their own context, for instance when a construct

ceases to be relevant over prolonged use. Last, 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.

An alternative approach for the measurement of the dynamics of experience over time relies on the elicitation of idiosyncratic self-reports of one's experiences with a product, i.e. *experience narratives*. Each narrative provides rich insights into a given experience and the context in which it takes place. Moreover, generalized knowledge may also be gained from these experience narratives. Such generalized knowledge may be reflected in 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 be improved to motivate prolonged use?

This manuscript makes two methodological contributions in this research problem. First, it highlights the labor-intensive nature of longitudinal studies, and proposes an alternative approach that relies on the elicitation of one's experiences with a product from memory. iScale, a tool designed with the aim of increasing users' effectiveness and reliability in recalling their experiences is theoretically grounded and empirically validated. Second, it proposes a computational approach that aims at supporting the researcher in the qualitative analysis of experience narratives. The proposed approach addresses two limitations of traditional qualitative analysis practices. First, qualitative analysis is a labor intensive activity which becomes increasingly a concern when qualitative data may be elicited from a large number of participants as in the case of iScale. Second, qualitative analysis has been shown to be prone to researcher bias as humans often rely on heuristics in forming judgments about the relevance or similarity of two or more data instances (Kahneman et al., 1982). The proposed approach aims at supporting the researcher through semi-automating the process of qualitative coding, but also minimizes the risks of overemphasizing interesting, but rare experiences that do not represent users' typical reactions to a product.

1.5 Manuscript Outline

We argue for a hybrid paradigm between reductionist and holistic approaches to evaluation. We propose two techniques, one grounded in *personal attribute judgments* and one in *experience narratives*. Both developed techniques aim at increasing the richness and diversity in obtained information while trying to create different levels of granularity of insight, thus enabling the researcher to move between abstracted, generalized insight and concrete, idiosyncratic and insightful information.

Part I

Understanding Interpersonal Diversity through Personal Attribute Judgments

Chapter 2 highlights the limitations of standardized psychometric scales and introduces personal attributes judgments. It introduces attribute elicitation techniques and in particular, the Repertory Grid Technique (RGT). It argues that the true value of RGT is in *quantifying rich qualitative insights* and highlights the limitations of relevant statistical techniques that are typically employed in the analysis of Repertory Grid data. It presents an initial Multi-Dimensional Scaling procedure that aims at identifying diverse views in Repertory Grid data. The procedure identifies distinct user groups in a sample population and derives a two-dimensional view for each respective user group. The technique is presented through a case study where users' views on a set of product concepts were contrasted to the ones of designers.

Chapter 3 presents a second Multi-Dimensional Scaling procedure that aims at identifying diverse views even within single individuals. The technique is applied on an existing dataset (Heidecker and Hassenzahl, 2007). It is illustrated that the -traditional- averaging analysis provides insight to only 1/6th of the total number of attributes in the example dataset. The proposed approach accounts for more than double the information obtained from the average model, and provides richer and semantically diverse views on the set of stimuli.

Part II

Understanding the Dynamics of Experience through Experience Narratives

Chapter 4 presents two studies that inquired into how users experiences with interactive products develop over time. In the first pilot study, grounded on reductionism, we asked participants to rate a novel product during the first week as well as after four weeks of use. In the second study six participants were followed after the purchase of a novel product and elicited rich experience narratives over a period of one month.

Chapter 5 presents iScale, a survey tool that aims at eliciting users' experiences with a product in the form of experience narratives. iScale employs sketching in imposing a process in the reconstruction of one's experiences from memory. The chapter motivates the development of two distinct versions of iScale which were grounded in two opposing theoretical approaches to reconstructing one's emotional experiences from memory. Finally, it presents two studies that compared the two different versions of iScale with traditional methods.

Chapter 6 proposes a semi-automated technique for the content analysis of experience narratives. The technique combines traditional qualitative coding procedures (Strauss and Corbin, 1998) with computational approaches for assessing the semantic similarity between documents (Salton et al., 1975). This results in an iterative process of qualitative coding and visualization of insights which enables to move quickly between high-level generalized knowledge and concrete and idiosyncratic insights.

Chapter 7 concludes the research described in this manuscript by reflecting on its contributions and sketching directions for future research.