

CAPTURING USER EXPERIENCES OF MOBILE INFORMATION TECHNOLOGY WITH THE REPERTORY GRID TECHNIQUE

Daniel Fallman

*Interactive Institute Umeå and
Department of Informatics
Umeå University, School of Architecture
Sweden*

John Waterworth

*Department of Informatics
Umeå University
Sweden*

Abstract: *We describe the application of the repertory grid technique (RGT) as a tool for capturing the user experience of technological artifacts. In noting the artificiality of assessing the emotional impact of interactive artifacts in isolation from cognitive judgments, we argue that HCI techniques must provide practical solutions regarding how to assess the holistic meaning of users' interactive experiences. RGT is a candidate for this role. This paper takes the reader step by step through setting up, conducting, and analyzing a RGT study. RGT is a technique on the border between qualitative and quantitative research, unique in that it respects the wholeness of cognition and does not separate the intellectual from the emotional aspects of the user experience. Compared to existing methods in HCI, RGT has the advantage of treating experiences holistically, while also providing a degree of quantitative precision and generalizability in their capture.*

Keywords: *user experiences, mobile HCI, repertory grid, design.*

INTRODUCTION

Adopted from the cognitive psychology of the 1970s and in force until relatively recently, the main theoretical approach to understanding human–computer interaction (HCI) was to view a person interacting with a computer generally as a disembodied information processor. Similarly, the standard methodological practice was to perform various lab-based quantitative experiments to gain empirical insight into the usability of a particular interactive device or environment, typically understood in terms of the specific qualities of the information processing involved. The nature of the users' *experiences* during interaction, that is, how he or she felt about it, was not considered or addressed.

In the last two decades, many of the limitations of this approach have been well documented within HCI by, for instance, Suchman (1987), Winograd and Flores (1986), Landauer (1991), and others. To a large segment of the HCI community, it has been clear for

many years that there is more to the interaction between human users and interactive artifacts than information processing, and that methods other than tightly controlled experiments are needed if more experiential aspects of interaction are to be captured. Thus, since the early 1990s, HCI researchers have increasingly explored broader issues to gain an understanding of the relationship between the user and the artifact in terms of, for instance, affective qualities, fun, and playability. In other words, researchers and practitioners are starting to consider the user not just as a processor of information and an experimental subject, but rather as an individual with hopes, desires, expectations, and emotions.

During the period when this change of perspective in HCI was gradually taking place, psychological approaches to cognition had already moved on. After a long period in the psychological wilderness, emotion became recognized within mainstream cognitive science as a fundamental component of cognition, of our making sense of the world. As neuroscientists such as Antonio Damasio (1994, 1999) pointed out, not only are our experiences limited without emotion, but we cannot make decisions. Affect is seen as an essential component of reasoning about the world, not an opposing force. Although we may loosely speak of emotion versus reason, both too much and too little emotion will have a negative impact on cognition, with the latter being the more pathological. Understanding the nature and varieties of conscious experience is also a central topic for contemporary cognitive science. For example, huge advances have been made in identifying the neural correlates of a range of subjective states, and relating these to verbal phenomenological reports and behaviors. Experiences and behaviors are viewed as two integrated effects of the same neural events, not as separate things.

In attempts to deal with and speak about these new issues in HCI, which are far more complex than the simple human processing and associated usability views they have come to replace, the concept of *user experience* has become a key concept in recent HCI research. While there is no unified theory about the role and implication of experience to design (Forlizzi & Battarbee, 2004), a number of efforts have been made recently within HCI to establish a better understanding of the role of user experience in interactive systems design (see, e.g., Fallman, 2003, 2006; Forlizzi & Battarbee, 2004; Forlizzi & Ford, 2000; Hassenzahl & Tractinsky, 2006; Ketola & Roto, 2008; Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009; McCarthy & Wright, 2004; Waterworth & Fallman, 2007).

A central issue in current user experience research is methodological: Exactly how do we best capture the experiences users have while being exposed to various designs? Purely quantitative measures, such as success rate and reaction time, do not seem to relate directly to users' experiences even though they may be useful in predicting some aspects of user performance under certain conditions. On the other hand, qualitative approaches, such as interviews and questionnaires, often lack any external validation and are limited in terms of generalizability and reliability. What is needed is a hybrid approach that provides a quantifiable and reliable measure, while also capturing subjective aspects of the experiences engendered by specific HCI designs.

In this paper, we provide an example of a candidate technique that we believe can be useful for getting insights into users' experiences of interactive artifacts in a quantitative way. We start from the position that interaction is about finding meaning, and that this involves judgments that result from a highly integrative blending of rational and affective elements, each relying on the other in producing a user's experience of an artifact. *Meaning* here refers to the sense individuals make of artifacts; we take things to mean what they are experienced

to be, reflecting the close coupling of rationality and affect. As observers of our own experiences, we cannot separate the two, except perhaps in extreme cases. In what follows, we describe and illustrate what we consider to be a promising technique for capturing the dimensions of meaning that characterize user experiences of technology in a holistic, yet also quantitative, way: the repertory grid technique (RGT).

THE REPERTORY GRID TECHNIQUE

The repertory grid technique (RGT) is a structured procedure of eliciting a repertoire of conceptual structures and for investigating and exploring them and their interrelations (Bannister & Fransella, 1985; Dalton & Dunnet, 1992; Landfield & Leitner, 1980). It has been found to be a useful technique for eliciting meaning in several different domains, for instance in organizational management, education, clinical psychology, and particularly in the development of knowledge-based systems (Boose & Gaines, 1988; Shaw, 1980; Shaw & Gaines, 1983, 1987).

RGT is a methodological extension of Kelly's (1955) personal construct theory. Kelly argued that we make sense of our world through our own construing of it. That is, we tend to model what we find in the world according to a number of personal constructs that are bipolar in nature and structure our experiences of the world. For instance, according to Kelly, we judge other people through forming personal constructs such as tall–short, light–heavy, handsome–ugly, and so on. A *construct* is essentially a single dimension of meaning for a person allowing two phenomena to be seen as similar and thereby as different from a third (Bannister & Fransella, 1985). Experiences arise from the interaction of multiple personal constructs.

What is a Repertory Grid?

While RGT is a technique for eliciting personal constructs, and a *repertory grid* is the outcome of a successful application of the technique. It is a table, a matrix, whose rows contain constructs and whose columns represent the *elements* of the phenomena under investigation. Repertory grids also typically embody a rating system used to relate each element quantitatively in relation to the qualitative constructs. An individual repertory grid table is constructed for each subject participating in a RGT study. This construction process, which will be described in detail later in this paper, is fairly straightforward. First, an individual participating in an elicitation session produces her (usage intended to be inclusive) own constructs, that is, what bipolar dimensions of meaning the person sees as the most important for talking about the elements (the investigated phenomena). The construct elicitation process is typically facilitated by the use of *triads*, through which the participant becomes exposed to sets of three elements at a time and is asked to describe and put a label on what he or she sees as separating one of the elements in the group from the other two. Second, after having provided her own individual, qualitative constructs, the participant is asked to rate the degree to which each element in the study relates to each bipolar construct according to some scale (typically a binary or Likert-type scale). Hence, in RGT, constructs and elements are the two building blocks of each individual's unique repertory grid table, and which are quantitatively related to each other by the use of some rating system. The constructs represent the qualities the

participants use to describe the elements in their own personal words (Fransella & Bannister, 1977). Constructs thus embody the participant's meaning and experience in relation to the study's elements.

RGT in Human–Computer Interaction

RGT has been found to be a useful technique for eliciting people's experiences and meaning structures in several different domains, including information systems (Tan & Hunter, 2002), education, clinical psychology, and particularly the development of knowledge-based systems (Boose & Gaines, 1988; Shaw, 1980; Shaw & Gaines, 1983, 1987). Despite its popularity in these fields, the interest in RGT from an HCI perspective peaked in the 1980s, with a special issue devoted to the topic in the *International Journal of Man-Machine Studies* in 1980. Since then, the technique's appearance in HCI-related literature has been sparse, while not completely nonexistent (see, e.g., Dillon & McNight, 1990; Grose, Forsythe, & Ratner, 1998; Hassenzahl & Wessler, 2000; Tomico, Karapanos, Levy, Mizutani, & Yamanaka, 2009). This lack of popularity may be due to fairly strong association with artificial intelligence and expert systems development in the 1980s, developments that came to epitomize the cognitivist viewpoint from which many HCI researchers were intent on distancing themselves.

Tan & Hunter (2002) recommend RGT as a means of studying the cognition of professionals and users of information systems in organizational settings, and review four examples of previous work focusing on its use for knowledge modeling. The emphasis of this kind of work is more on identifying experts' cognitive rules than on the nature of subjective experiences with technology. But recently, there has been a modest resurgence of interest in RGT as a means of capturing dimensions of user experiences with technology, as shown in research on loudspeaker array design (Berg, 2002) and subjective aspects of immersive virtual reality (Steed & McDonnell, 2003); and, more recently, to help understand cross-cultural differences in the experience of different designs of writing pen (Tomico et al., 2009).

The intention of the present paper is to further explore the potential of RGT, and to bring it to the attention of the HCI community as a possible integrative approach to understanding user experiences in HCI. This approach assumes that emotion and reason are essential and interrelated parts of making sense of the world, and provides results that are both subjective and quantitative. The following sections take the reader step by step through the setting up and carrying out of an HCI study using RGT in the context of mobile interaction devices.

USING RGT TO CAPTURE THE EXPERIENCE OF USING MOBILE INFORMATION TECHNOLOGY

In the study described below, we were interested in how people experience mobile information technology, as embodied in existing products and newly developed research prototypes. In addition to a general interest in how people relate to this kind of technology, we wanted particularly to gain empirical insight into what kinds of meanings people ascribed to the different styles of interaction these various devices embodied. The study involved existing off-the-shelf devices, as well as a number of research prototypes that represent a range of alternative means of interaction.

Participants

The empirical data collection process was carried out over a period of 3 weeks. In total, 18 participants took part in the study, all of which had previously volunteered by signing up for a scheduled time slot. Of the total number of participants, 14 (78%) were males and 4 (22%) were females. Eight of the participants (44%) were in the age span of 20–29, seven (39%) were 30–39 years of age, two (11%) were 40–49, and one (6%) was 50–59 years. As assessed by a preparatory questionnaire, three participants (16%) rated themselves as 3 on a 5-graded scale of self-estimated computer literacy, 14 (78%) rated themselves 4, while only one (6%) indicated 5. On a similar scale from 1 to 5, when asked to rate their previous exposure to mobile information technology, one participant (6%) responded with a 2, six (33%) rated themselves as 3, nine participants (50%) rated themselves 4, while two (11%) considered themselves to be 5 out of 5. As a sign of appreciation for their participation in the study, participants were provided cinema tickets. Each session lasted from 45 minutes to two hours, averaging slightly more than an hour. All participants took part in the study individually, with only the participant and the experimenter in the room. With the exception of a single native English speaker, the other 17 participants were native Swedish speakers. The study was carried out in each participant's native language and carefully translated for this paper.

Step 1: Element Familiarization

All 18 sessions began with the participant being exposed to seven different mobile information technology devices. Three of them were examples of existing devices; a Compaq iPaq H3660 personal digital assistant (PDA, known in the study as E0), a Canon Digital Ixus 300 digital camera (E1); and a Sony Ericsson T68i mobile phone (E2).

Four research prototypes were also part of the study (see Figure 1, a–d). The ABB Mobile Service Technician (E5, Figure 1a) is a wearable support tool for service technicians in vehicle manufacturing (Fallman, 2002). The Dupliance prototype (E4, Figure 1b) is a physical/virtual communication device for preschool aged children (Fallman, Andersson, & Johansson, 2001). The Slide Scroller (E3, Figure 1c) combines a PDA with an optical mouse to form a novel way of interacting with Web pages on palmtop-size displays (Fallman, Lund, & Wiberg, 2004). Finally, the Reality Helmet (E6, Figure 1d) is a wearable interactive experience that alters its user's perceptual experience (Fallman, Jalkanen, Lörstad, Waterworth, & Westling, 2003; Waterworth & Fallman, 2003).

Each session started with the seven devices being presented, one by one, to the participant. We provided brief (3–5 minutes each) introductions to the different contexts of the four research prototypes and the projects from which they originated. The participant was then able to try out each device for as long as necessary in order to become familiar with it. The session organizer was always available during the session and willing to answer any questions posed by the participants.

Step 2: Construct Elicitation

After the preparatory questionnaire had been completed, the elicitation of a participant's constructs for the seven elements (devices) began. Each participant sat at a table opposite to the

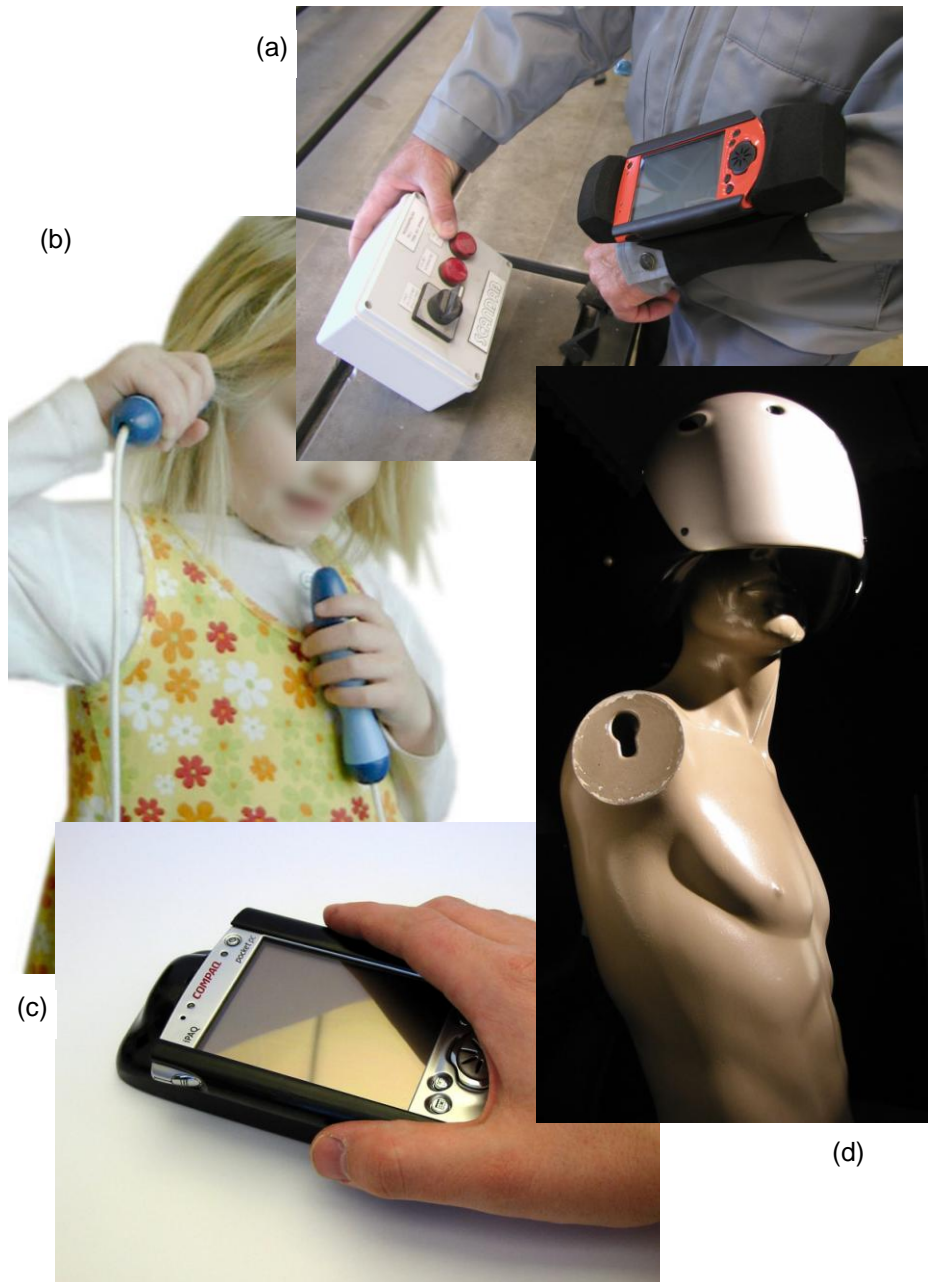


Figure 1. The four research prototypes that, together with three existing devices, were part of this study.

experimenter. On the table, seven palm-sized cards were displayed. Every card contained the following: a photograph of one of the devices; a label on which the name of the device was printed; and the identification number used for organizing the study (i.e., E0 to E6). In each session, the participant was exposed to the seven devices in groups of three; this is known as *triading* in RGT's technical language. Each triad was chosen from a list randomized prior to the study.

On a paper-based form designed especially for this study, the experimenter put down three identification numbers taken from a pre-prepared list, for instance E0, E4, and E5. The experimenter and the participant then together found the corresponding cards on the table and grouped them in front of the participant, while the remaining four cards were put aside. The participant was then asked to think of a property or quality that she considered notable enough to single out one of the three elements (devices) in the triad, and to put a name or label to that property. For instance, among a group of E1, E2, and E3, Participant 10 singled out E1, and labeled her experience as “warm.” The participant was then asked to put a name or label on the property or quality that the other two devices in the triad shared in relation to the experience of E1. Participant 10 decided to collectively label E2’s and E3’s shared quality as “cold.”

Some of the participants were fairly quick in finding what they saw as appropriate labels to put on their experiences; others could remain silent for quite some time, thinking carefully to themselves, while a few others discussed loudly and in detail their thoughts and ideas with the experimenter. Although the experimenters tried to answer questions and generally took part in discussions initiated by the participants, we were careful not to generate or imply properties or concepts, in order to avoid putting our words into the participant’s mouth. To be able to keep the relation between construct and originator throughout the study, the suffix (S10) was added to each construct elicited from Participant 10. Hence, in this case, the elicited personal construct was Warm (S10)–Cold (S10).

On the form there was also a preprinted table containing the elements, each with its own 7-grade Likert-type scale. After the triading session, the form was handed over to the participant with the instruction to grade each of the seven elements according to the bipolar scale that had just been constructed from the participant’s own concepts. That is, for each element of the study as a whole—including those that did not appear in the specific triad from which a particular construct pair was established—the participant was asked to rate or grade that element on a 7-point scale, 1 would represent a *high degree of the property found to be embodied in a singled out device* (e.g., in the case of Participant 10, “warm”), 7 would represent a *high degree of the property embodied by the two other devices in the specific triad* (i.e., “cold”).

The Likert scale is the most widely used scale in survey research for measuring attitudes in which respondents are asked to express their strength of agreement, typically using an odd number of response options. For this study, we chose to apply a 7-grade scale, for two primary reasons. First, compared to an even-grade (a so-called forced choice scale), a scale with an odd number of choices does not force people to make choices that might not reflect their true positions). A grade 4 out of 7 thus indicates, statistically, that a construct has no particular meaning for a given element. This is important since the constructs in a repertory grid are constructed from triads in which only three out of seven elements appear. Second, because some people do not like making extreme choices (i.e., 1 or 7 out of 7), the 7-grade scale provides richer data than, for instance, 3- or 5-grade scales.

Thus, for each triad exposed to a participant, two kinds of data were collected. First, a personal construct was elicited (i.e., a one-dimensional semantic space that the participant thought meaningful and important for discussing and differentiating between the elements of a triad). This process provided the study with qualitative data: insight into the participant’s own meaning structures, values, and preferences. Second, since each elicited personal bipolar construct was then used as the scale by which the participant rated all seven elements in the study using a 7-point Likert scale, data were also gathered about the degree to which

participants thought their construct had relevance to a specific element. This provided the study with quantitative data used to find out how the different elements compare and relate to each other and to the constructs, described in detail below. This analysis reveals, or at least suggests, whether or not, for example, Participant 10's construct "warm-cold" is purely literal (i.e., referring to the actual temperature of the artifact) or metaphorical (i.e., referring to the emotional effect the artifact has on the participant). The same kind of statistical analysis would not have been possible if we had asked the participants to rank rather than rate the elements.

To keep the length of the sessions roughly equal and in order not to make our participants weary, we decided to limit each session to 10 triads. Thus, from the 18 participants we elicited 180 pairs of personal constructs (i.e., 360 different concepts the participants thought meaningful and relevant) for describing their experiences of mobile information technology. At this point, it should be noted that a specific advantage of the RGT approach is that it is not necessary for the experimenter to share the specific meaning structures a participant holds in relation to an elicited construct at the time of elicitation. These are revealed during analysis by comparing the data connected with elicited constructs to data connected with other groups of elicited constructs.

ANALYSIS OF REPERTORY GRID DATA

While RGT is an open approach that results in a number of highly individual repertory grid tables, some basic structures are shared among the participants. Each table in this study consisted of a number of bipolar constructs; a fixed number of elements (7); and a shared rating system (a scale of 1 to 7). From this setup, there are at least two basic ways in which different people's repertory grid tables may be compared and analyzed interpersonally (i.e., to compare different people's repertory grids in different ways).

First, the finite number of elements and the shared rating system provide the basis for applying statistical methods that search for variations, similarities, and other kinds of patterns in the series of numbers occurring in the numerical data (the ratings). Using relational statistical methods, it becomes possible to compare and divide all constructs from all participants into groups of constructs showing some degree of similarity. This may result in interesting and unexpected correlations between constructs whose relation would most likely have remained unnoticed if one were only looking for semantic similarity. This method may hence be called *semantically blind*, since it is driven primarily by each construct pair's quantitative data in relation to elements.

Second, several seemingly semantically related and overlapping groups of construct pairs appeared across the study's participants. Some similar bipolar scales, for instance, *young-old*, *appliance-multifunctional*, and *work-leisure*, can be spotted among the responses from several of the participants. It would be possible to go through the list of all participants' constructs and gather in groups those that bear semantic resemblance to each other, and analyze these groups (e.g., using discourse analysis). This approach could be regarded as *statistically blind*, since it is driven by an interpretation of the semantic content of the constructs, not taking the numerical ratings into account.

Both of these approaches would result in a number of groups of constructs. In this particular study, we were primarily interested in finding correlations between different

constructs that may or may not seem by semantic resemblance to belong together, but which according to their ratings do. From this, it appeared that a semantically blind statistical approach that compares ratings would be the best choice for exploring the data set.

Step 3: Participant-Level Analysis

The manually collected data from the 18 participants was compiled and put into the WebGrid-III application, a frequently used and feature-rich tool for collecting, storing, analyzing, and visually representing repertory grid data (Gaines & Shaw, 1980, 1993, 1995). Each participant's repertory grid table was used as the basis for three different ways of presenting the data graphically, increasingly driven by and dependent on statistical methods of analysis.

First, a Display Matrix was generated. As the most basic way of presenting a repertory grid, this table simply lays out the numerical results of all constructs for all elements. Second, a FOCUS Graph was constructed for each participant. Here, both elements and constructs are sorted using the FOCUS algorithm (Gaines & Shaw, 1993, 1995; Hassenzahl & Wessler, 2000) so that similar ones are grouped together.

Third, the PRINCOM Map provides principal component analysis of the repertory grid data. The grid is rotated and visualized in a vector space to facilitate maximum separation of elements in two dimensions (Gaines & Shaw, 1980; Slater, 1976). For more detailed information and discussion about these common ways of analyzing and visualizing repertory grid data, see Gaines & Shaw (1993, 1995), Shaw (1980), and Shaw & Gaines (1998).

Step 4: Statistical Analysis of Multiparticipant Data

For our study, we were interesting in seeing if any patterns or other kinds of relationships between different participants' repertory grids could be derived. But how could these highly individual and subjective personal constructs be compared with each other in practice? To be able to perform statistical analysis on multiparticipant data, all 180 bipolar constructs of the participants were put into the same, very large repertory grid. This huge grid then became subject to various kinds of analyses similar to those applied to each individual participant's repertory grid. Hence, a DISPLAY matrix, a FOCUS graph, and a PRINCOM map were constructed from the WebGrid-III application using all the data. These diagrams are immense and unstructured, so the task at this point became to refine and bring order into the data set.

Statistical analysis may be performed on repertory grid data to find similarities and other kinds of patterns among the constructs elicited from different participants. Finding constructs that share a rating pattern indicate that they, mathematically, belong to the same group. This suggests that the coherence in rating also reflects coherence in experience, but one which may have been expressed differently in the semantic terms used. A group whose constructs share a unique topology in ratings thus becomes seen as a specific dimension of meaning in relation to the elements of the study. The part played by the researcher in this process is, through semantic analysis of the constructs that make up such groups, to establish what conceptual similarity they share.

Finding Groups by FOCUS Analysis of Data (1st Round)

To discover groups within the data set, the large repertory grid constructed from all the participants' individual grids was subject to two cycles of FOCUS clustering. The difference between the two rounds was in the manipulation of two rules that were applied to distinguish groups or clusters in the data.

The first rule was that the threshold level for regarding two constructs as similar was placed at 90%, that is, the constructs needed to share at least a 90% consistency in rating to be grouped together. Naturally, this rule may be discussed and questioned in a number of ways. Most obviously so, why was the 90% mark designated? In reality, this analysis effort most often needs to iterate a few times with different percentages in order to get to know the data set. Settling with 90% as a first rule of the first round was aimed at keeping a balance between (a) the number of clusters that emerge, (b) the size of these clusters, and (c) a reasonable level of internal coherence within each cluster. A higher threshold higher, say at 95%, generates clusters with a stronger degree of internal consistency, but they also become quite few in number. In addition, each cluster becomes fairly limited in terms of the number of contributing constructs. Using an overly high threshold also would leave out many of the constructs from the study and much of the study's semantic "flesh"—the place where the participants' meanings and experiences reside—would be lost. On the other hand, an overly low threshold, set at 60% or 70%, would result in almost all constructs being part of a cluster—thus embracing the lion's share of the meanings with which the participants have charged the elements—but these clusters would be very large in terms of number of constructs, and thus decreasing the clarity or definement of the element they represent. And, since each cluster would consist of a large number of constructs, a low threshold would also result in a small number of clusters in total. Thus, an overly low threshold would associate a particular construct with too many of the other constructs, where meaning would disappear in a few, large, and unmanageable clusters. Through the exploration of different threshold levels during this round, a threshold of 90% was found to be reasonable for a first statistical clustering of the constructs.

As a second rule of the first round, a *cluster* was defined as consisting of three or more constructs. When applying these two rules on the data set, 17 groups emerged consisting of 3 to 12 constructs. Each group was named with the prefix *A* followed by the group's number from top to bottom on the chart generated by the FOCUS algorithm.

Finding Groups by FOCUS Analysis of Data (2nd Round)

While the first round provided a number of statistically coherent groups, a large number of the grid's constructs had not been included. The purpose of the second round was to manipulate the rules for forming clusters so that more of the participants' constructs were included, even at the cost of lower internal coherence. This was done by lowering the threshold level to 85%, so that larger clusters developed around those established in Round 1, as well as a number of completely new clusters. To counterbalance the weaker internal coherence in rating these clusters, the second rule was made more stringent by the additional rule that clusters in this round needed to be made up of four or more constructs. Each of these groups was then named with the prefix *B* and the group's number. Twelve groups were established in this round.

Step 5: Naming Groups by Semantic Analysis

The groups identified so far may be regarded as representing the 29 most pertinent dimensions of the participants' understandings of the elements of the study. The first task of the next step was to create 29 new repertory grids based on the contributing constructs of a group. A Display Matrix, a FOCUS graph, and a PRINCOM map were also generated for each group. The analysis, up to this point, had remained statistical rather than semantic: Each of the 29 groups consisted of a number of constructs whose *ratings* grouped them together. But to be able to address a specific group as a shared bipolar concept, an interpretative analysis became necessary. Each dimension of each construct in each group was thus carefully semantically reviewed and interpreted, and one—or, if needed to better capture the character of the cluster, two or three—of the existing labels (from different participants) was chosen to characterize the group as a whole, and used to form a new bipolar construct representing the group.

At least two issues need to be highlighted in relation to this activity. First, not all constructs in a group fit perfectly well with each other semantically. Some constructs are also odd, unusual, and obviously point at something else than most others in the group. While this is not uncommon when dealing with large amounts of quantitative data, it puts the researcher in the uncomfortable position of having to make judgments about which constructs to include in a group and which to disregard in order to capture the general tendency of the group. In a few cases, no semantic resemblance and no recognized meaning structure could be established from the particular constructs of the group in question, and these groups were excluded at this stage in the procedure. In addition, some of the groups at the B-level are formed around A-level clusters, where the broadening has not always been found to provide any richer semantic information than their corresponding groups at the A-level. Thus, six B-level groups were excluded.

Additionally, even though the interpretative nature of this labeling means that the following analysis is not completely data driven, the potential hazards of experimenter biases and pure misunderstandings are reduced by choosing from existing participants' labels to capture the character of a group, rather than creating new ones. As an example of how this labeling was carried out, the group A16 consists of three contributing constructs, with “Cosmetical (S18),” “Consumer product (S14),” and “Device (S1)” on one end and “Mechanical (S18),” “Professional product (S14),” and “Tool (S1)” on the other. Here, *Device* (A16) was chosen to represent the former and *Professional tool* (A16) to represent the latter end.

Step 6: Calculating Mean and Median Ratings

If these groups, with their labels as representatives, are treated as constructs, it is possible to form a new repertory grid consisting of these 23 groups/constructs and the original elements. But to be able to statistically analyze how they relate to each other and to the elements of the study, a rating for each construct on each element needs to be incorporated into the new repertory grid table. Rather than using the arithmetic mean, these calculations relied on the median value. This was found to provide a result that seems more true to the rating of the participants, one in which the influence of single, extreme values at odds with the majority of the values in the group was de-emphasized. For each value, a standard deviation was calculated, providing clues to which values in a group are the most uncertain. Comparing the

standard deviations for the ratings across the elements of a group, as well as the value for the average absolute deviation from median, tells us something about how certain a specific rating is and provides clues to the lack of agreement by the participants on specific elements.

Step 7: Interpreting and Presenting the Result

When applying an 85% threshold to these 23 clusters and their ratings, the FOCUS algorithm further partitioned them into three groups of four or more constructs, as well as a single clustering between two additional constructs. These clusters may again be treated as groups, and hence, given these four new constructs formed from these clusters together with the remaining six non-clustered constructs, the statistical analysis leaves us with not 23 but rather 10 unique dimensions of the way in which the participants have experienced the devices of mobile information technology that were part of the study.

These 10 dimensions are presented as a FOCUS graph (Figure 2) and as a PRINCOM map (Figure 3), which also shows how the different elements relate to each other. The FOCUS graph sorts the grid for proximity between similar elements and similar constructs while the PRINCOM map makes use of principal component analysis to represent the grid in minimum dimensions (Shaw & Gaines, 1995). These 10 dimensions are thus the most significant ways in which the participants experienced the elements of the study.

The results give us a graphic account of how participants construed the seven devices and, in particular, how their experience of each related to that of the others. We must be cautious of using the construct labels literally, but it is clear that the Reality Helmet, as an example, is semantically distant from the digital camera, as shown by their opposing positions

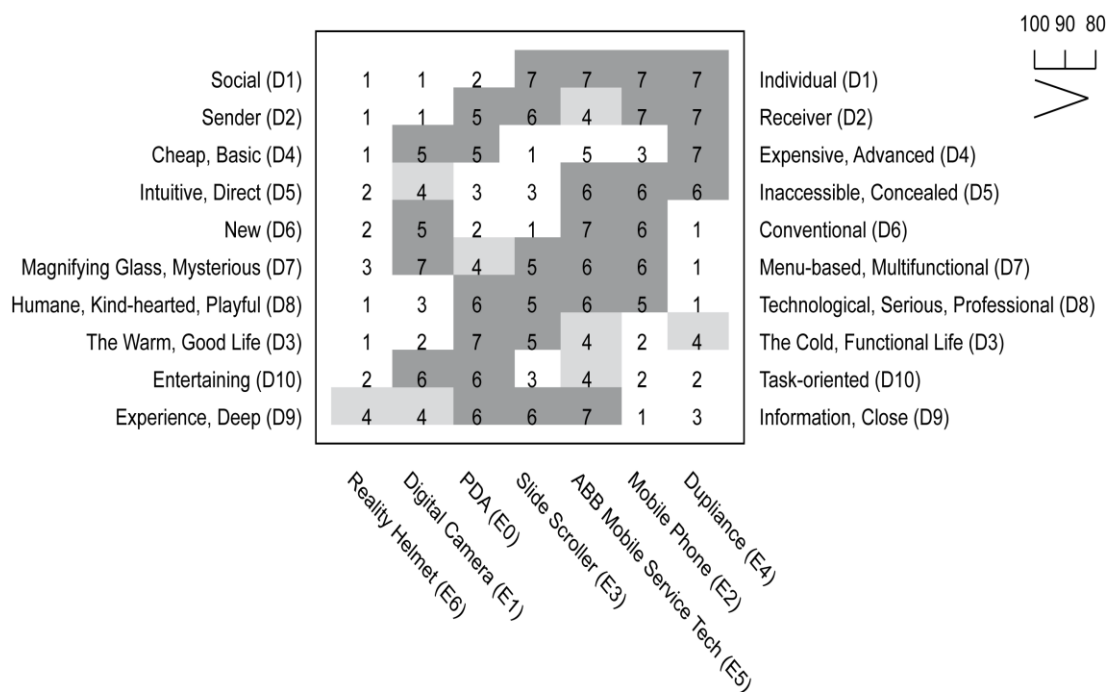


Figure 2. The resulting 10 unique dimensions (D) of the study presented as a FOCUS graph.

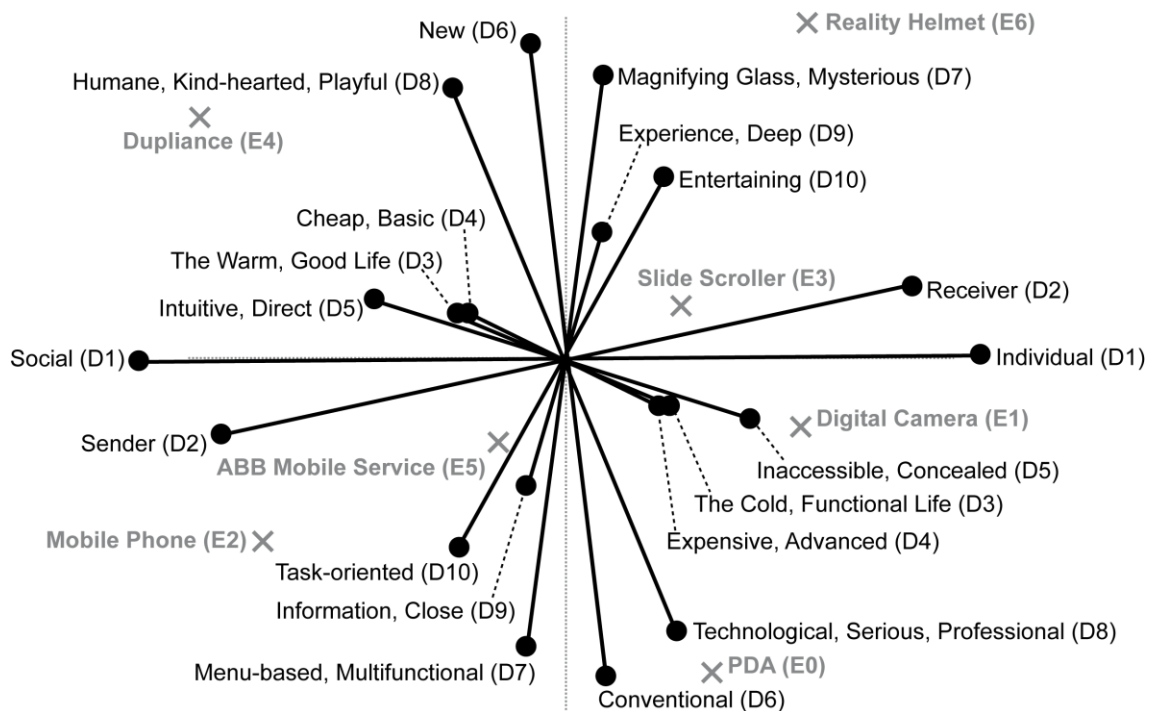


Figure 3. The 10 dimensions presented as a PRINCOM map.

on dimensions such as task-oriented (Digital Camera) versus entertaining (Reality Helmet). Several of the devices were experienced as relatively “social” (Dupliance, Mobile Service Technician, Mobile Phone) as compared to others that were more “individual” (Reality Helmet, Slide Scroller, Digital Camera, PDA). The Dupliance was associated with positive attributes such as “humane,” “warm,” and “intuitive,” whereas the Digital Camera was seen as more “cold” and “concealed.” The Mobile Service Technician and the Mobile Phone were quite close to each other, and both were associated with “task-oriented.” Taken as a whole, the dimensions provide a wealth of information about how these users experienced the seven artifacts, and how they compared with each other.

DISCUSSION

This paper is primarily concerned with the use of RGT as a methodological tool for getting at people’s experiences of using technology, relevant to the current concerns of HCI. We have shown how the procedure may be used to assess the experiences people have of designs, as in the study described above. In the following sections, we reflect further on the use of RGT as an element in research and design efforts, spotlighting ways in which it differs from other approaches in HCI.

Moreover, we point out that the RGT also can be employed during design, when included as a part of an iterative design cycle that aims for the user to have certain experiences. We might want to design, for example, a device that is experienced in a similar way to another existing device. This point is taken up in the concluding section of the paper.

RGT is an Open Approach

There are arguably some potential advantages of using RGT as compared to other candidate techniques for gaining insight into people’s meaning structures. While RGT is a theoretically grounded, structured, and empirical approach, it is not restricted or limited to already existing, preprepared, or researcher-generated categories. Alternative approaches showing the same kind of openness as RGT include the semantic differential, discourse analysis, ethnography and similar observational methods, and unstructured interviews.

RGT is both Qualitative and Quantitative

Because a repertory grid consists of not only the personal constructs themselves but also a rating of them in relation to other elements in the study, the researcher not only gains insight into which are the meaningful constructs, but also the degree to which a particular construct applies or does not apply to a particular element. Hence, the RGT technique perhaps may be characterized best as being on the border between qualitative and quantitative research: a hybrid, “quali-quantitative” approach (Tomico et al., 2009).

On the one hand, a repertory grid models the individual perspectives of the participants, where the elicited constructs represent the participants’ subjective differentiations. It may be used as such for various kinds of interpretative semantic analysis. On the other hand, since systematic ratings of all elements on all constructs result in a repertory grid consisting not only of elements and constructs but also of quantitative ratings, the resulting repertory grid may be subject to different kinds of quantitative analyses as well. The quantitative aspect of the RGT also provides the necessary means for comparing participants’ grids with each other, using contemporary relational statistical methods. While RGT is reliant on statistical methods, semantic interpretation is sometimes needed to carry out specific parts of the analysis. By consistently using codes and markers, it is possible to track these interpretations back to the original data set.

RGT Results are Relational Rather than Absolute

Because RGT relies on comparisons between different elements, all results—such as the 10 unique dimensions of the example study—should be regarded as relative to the group of elements included in the study. The outcome of a study using this technique is not a set of absolute values. Rather, studies using RGT produce insights into people’s experiences of particular things and the relationships between them. This potential disadvantage of the method was addressed in our example study by including already existing mobile information technology devices in the study to which the new research prototypes can be related. Doing so provided a result that, while still not absolute, nevertheless has become *situated*. In this respect, use of RGT is similar to the application of psychophysical rating scales to capture observers’ perceptual judgments, which are always relative to the range of stimuli presented (e.g., Helson, 1964, Poulton, 1989; Schifferstein, 1995). Experiences can never be captured with the absolute precision of some physical measurements. Experiences can only ever be judged relative to other experiences, and the RGT approach emphasizes this fact.

RGT Addresses the User's Experience Rather than the Experimenter's

A famous contemporary and contrasting attempt at identifying and quantifying meanings and attitudes comes from the work of Charles Osgood in the 1950s (Osgood, Suci, & Tannenbaum, 1957). His semantic differential technique was developed to let people give responses to pairs of bipolar adjectives in relation to concepts presented to them (Gable & Wolf, 1993). The main adjectives used by Osgood included evaluative factors (e.g., good—bad), potency factors (e.g., strong—weak), and activity factors (e.g., active—passive). Each bipolar pair hence conceptually suggests a one-dimensional semantic space, a scale on which the participant was asked to rate a concept. Given a number of such pairs, the researcher is able to collect a multidimensional geometric space from every participant, much like the RGT approach.

However, researchers have raised a number of objections to and reservations about Osgood's technique. Among the most important is the recognition that the technique seems to assume that the adjectives chosen by the experimenter have the same meaning for everyone participating in the study. Also, since the experimenter provides the participants with the bipolar constructs, the former tends to set the stage, that is, provides the basic semantic space, for what kinds of meanings the participant can express for a particular concept. When participants merely rate construct pairs given to them, they are able to dismiss certain pairs as not appropriate or of no significance for a particular concept, but they have no way of suggesting new adjectives that they feel are more appropriate for describing something.

In contrast, the RGT approach does not impose the experimenter's constructs on participants. Rather, the method aims to elicit the users' own understanding of their experiences. In its first phase, RGT is clearly focused on eliciting constructs that are meaningful to the participant, not to the experimenter. The data in a particular participant's repertory grid is not interpreted in the light of the researcher's own meaning constructs.

Invested Effort

One disadvantage of RGT is that it requires a substantial investment of effort by both the experimenter and the participants at the time of construct elicitation, as compared to most quantitative methods. This has implications for both how many participants it is reasonable to have in a study, as well as for the length of each eliciting session. Although it would be better to expose each subject to as many triads as possible, doing so would not have been practically viable in this study, for the following reasons.

First, from around triad 8, we noticed that most participants' ability to find meaningful construct pairs began to decrease significantly, which was something that many of the participants also stated explicitly. Second, 10 triads also kept the length of each session to slightly more than an hour on average, which seemed to be a reasonable amount of time to expect people to concentrate on this kind of task.

Third, with seven elements, the number of possible unique triads exceeds 40, which is clearly far too many to expose to each participant (at least, if there is only a movie ticket at stake). This means that each participant was only exposed to a subset of all possible combinations of triads. However, because different participants were exposed to different triads, each unique group has been covered in the study as a whole.

On the other hand, RGT is more efficient and less time-consuming than most other fully open approaches, such as unstructured interviews and explorative ethnography. And, because the personal constructs elicited from participants constitute the study's data, it follows that using the RGT significantly reduces the amount of data that needs to be analyzed, compared with transcribing and analyzing unstructured interviews or ethnographic records.

Specific Issues Regarding the Elicitation Process

Two potential problems regard the actual conduct of constructing repertory grids. While these are generally not unique to RGT, they are worth noting. First, for various reasons, participants may feel inclined to provide the experimenter with socially desirable responses. In other words, a participant may experience a sense of social pressure during the elicitation session that makes her try to give the experimenter the “right answer.” Second, some participants may, again for various reasons (e.g., that they feel uncomfortable in the situation, do not really have the time for the session, do not want to or cannot concentrate, do not really understand the purpose or doubt the study's usefulness, etc.), come to develop a habit of consistently providing moderate answers, or always either fully agreeing or disagreeing with their own constructs.

CONCLUSIONS

In this paper we have commented on the artificiality of assessing the emotional impact of interactive artifacts in isolation from cognitive judgments. We stressed that both emotion and reason are inherently part of any cognitive appraisal, and underlie the user's experience of an artifact. We suggested that studying the one without the other is – literally – meaningless. What HCI needs are techniques that recognize this and that provide practical solutions to the problem of how to assess the holistic meaning of users' interactive experiences.

In this light, a candidate method, the repertory grid technique (RGT), may partly fill this need, and has been presented, discussed, empirically exemplified, and explored. RGT was found to be an open and dynamic technique for qualitatively eliciting people's experiences and meanings in relation to technological artifacts, while at the same time providing the possibility for data to be subjected to modern methods of statistical analysis. The RGT may as such best be described as a research method on the border between qualitative and quantitative research. An example from the area of mobile HCI was used to take the reader step by step through the setting up, conducting, and analyzing of an RGT study.

How should a designer of interactive experiences think about the 10 dimensions of mobile technologies found in this study? Are they only relevant to this study and these devices, or are they general enough to provide a sound understanding of users' experience mobile information technology? The answer probably lies somewhere between these two possibilities.

Since RGT relies on comparisons between different elements, all results—such as the 10 unique dimensions surfaced in this study—must be regarded as relative to the group of elements that were included in the study. The 10 dimensions speak of something that is specifically about the seven technology designs provided to the participants. In a statistical sense, the resulting dimensions are relational to these seven devices. There is no way of

knowing whether they would change dramatically if an eighth device were to be added, without doing such an extended study.

But this limitation was to some extent addressed in the study by including already existing mobile information technology devices to which the new research prototypes can be related. Doing so provided a result that, while still not absolute, nevertheless has become more situated. It would not do justice to the study and the effort put into it by the participants to argue that the results are only valid within the study itself. On the contrary, we believe that the results from this study and the approach it illustrated could be useful for designers of mobile information technology, not the least as a tool for design.

Given that a team of designers wants to provide form and content to a mobile device that should embody certain characteristics, there are at least two ways in which this study can be used to guide the process. First, they may take the three existing devices as a basis and consider the four prototypes to provide a large number of alternative design dimensions. If they want their design to provide its users with a sense of mysteriousness, for instance, then aspects of the Reality Helmet may be taken as influence. Second, designers may use this study as the basis for designing and conducting their own studies in similar ways. If they want to find out whether their design really is experienced as mysterious, they can set up and conduct their own repertory grid study in a similar fashion, perhaps even using the same existing devices as were used here. Such comparisons can at least provide some hints and traces of meaning that may be very useful for further design work. The design team may also wish to embed small repertory grid studies throughout the production cycle to monitor designs against some sought-after set of qualities of user experience: These grids could become a recurring element in organizing the process of interactive artifact design.

RGT is unique in that it respects the wholeness of cognition: It does not separate the intellectual from the emotional aspects of experiences. At the same time, it acknowledges that each individual creates her own meaning in the way she construes things to be, in the context in which they are experienced. RGT has the advantage of treating experiences holistically, while also providing a degree of quantitative precision and generalizability in their capture.

REFERENCES

- Bannister, D., & Fransella, F. (1985). *Inquiring man* (3rd ed.). London: Routledge.
- Berg, J. (2002). Systematic evaluation of perceived spatial quality in surround sound systems. Doctoral Thesis. Luleå University of Technology (2002: 17), Sweden.
- Boose, J. H., & Gaines, B. R. (Eds.). (1988). *Knowledge acquisition tools for expert systems*. London: Academic Press.
- Dalton, P., & Dunnet, G. (1992). *A psychology for living: Personal construct psychology for professionals and clients*. London: Wiley.
- Damasio, A. (1994). *Descartes' error: Emotion, reason and the human brain*. New York: Penguin Putnam.
- Damasio, A. (1999). *The feeling of what happens: Body, emotion and the making of consciousness*. San Diego, CA, USA: Harcourt Brace and Co., Inc.
- Dillon, A., & McKnight, C. (1990). Towards a classification of text types: A repertory grid approach. *International Journal of Man-Machine Studies*, 33, 623–636.
- Fallman, D. (2002). Wear, point, and tilt: Designing support for mobile service and maintenance in industrial settings. In *Proceedings of DIS 2002: Designing Interactive Systems* (pp. 293–302). New York: ACM Press.

- Fallman, D. (2003). Design-oriented human-computer interaction. In *Proceedings of Conference on Human Factors in Computing Systems (CHI 2003)*; pp. 225–232). New York: ACM Press.
- Fallman, D., Jalkanen, K., Lörstad, H., Waterworth, J., & Westling, J. (2003). The Reality Helmet: A wearable interactive experience. In *Proceedings of SIGGRAPH 2003: International Conference on Computer Graphics and Interactive Techniques, Sketches & Applications* (p. 1). New York: ACM Press.
- Fallman, D., Lund, A., & Wiberg, M. (2004). ScrollPad: Tangible scrolling with mobile devices. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS '04; on CD)* Washington, DC: IEEE Computer Society
- Fallman, D. (2006, Nov). *Catching the interactive experience: Using the repertory grid technique for qualitative and quantitative insight into user experience*. Paper presented at Engage: Interaction, Art, and Audience Experience, Sydney, Australia.
- Fallman, D., Andersson, N., & Johansson, L. (2001, June). *Come together, right now, over me: Conceptual and tangible design of pleasurable dupliances for children*. Paper presented at the 1st International Conference on Affective Human Factors Design, Singapore.
- Forlizzi, J., & Ford, S. (2000). The building blocks of experience: An early framework for interaction designers. In *Proceedings of DIS 2000: Designing Interactive Systems* (pp. 419–423). New York: ACM Press.
- Forlizzi, J., & Battarbee, K. (2004). Understanding experience in interactive systems. In *Proceedings of DIS 2004: Designing Interactive Systems* (pp. 261–268). New York: ACM Press.
- Fransella, F., & Bannister, D. (1977). *A manual for repertory grid technique*. London: Academic Press.
- Gable, R. K., & Wolf, M. B. (1993). *Instrument development in the affective domain* (2nd ed.). Boston: Kluwer Academic Publishers.
- Gaines, B. R., & Shaw, M. L. G. (1980). New directions in the analysis and interactive elicitation of personal construct systems. *International Journal Man–Machine Studies*, 13, 81–116.
- Gaines, B. R., & Shaw, M. L. G. (1993). Eliciting knowledge and transferring it effectively to a knowledge-based system. *IEEE Transactions on Knowledge and Data Engineering*, 5(1), 4–14.
- Gaines, B. R., & Shaw, M. L. G. (1995). WebMap: Concept mapping on the web. *World Wide Web Journal*, 1, 171–183.
- Grose, M., Forsythe, C., & Ratner, J. (Eds.). (1998). *Human factors and web development*. Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Hassenzahl, M., & Wessler, R. (2000). Capturing design space from a user perspective: The repertory grid technique revisited. *International Journal of Human–Computer Interaction*, 12, 441–459.
- Hassenzahl, M., & Tractinsky, N. (2006). User experience: A research agenda [Editorial]. *Behavior & Information Technology*, 25, 91–97.
- Helson, H. H. (1964). *Adaptation-level theory*. New York: Harper & Row.
- Kelly, G. (1955). *The psychology of personal constructs* (Vols. 1 & 2). London: Routledge.
- Ketola, P., & Roto, V. (2008, June). *Exploring user experience measurement needs*. Paper presented at the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM), Reykjavik, Iceland.
- Landfield, A. W., & Leitner, L. (Eds.). (1980). *Personal construct psychology: Personality and psychotherapy*. New York: Wiley.
- Landauer, T. (1991). Let's get real: A position paper on the role of cognitive psychology in the design of humanly useful and usable systems. In J. M. Carroll (Ed.), *Designing interaction: Psychology at the human-computer interface* (pp. 60–73). New York: Cambridge University Press.
- Law, E., Roto, V., Hassenzahl, M., Vermeeren, A., & Kort, J. (2009). Understanding, scoping and defining user experience: A survey approach. In *Proceedings of Conference on Human Factors in Computing Systems (CHI 2009)*; pp. 719–728). New York: ACM Press.
- McCarthy, J., & Wright, P. (2004). *Technology as experience*. New York: The MIT Press.

- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Chicago: University of Illinois Press.
- Poulton, E. C. (1989). *Bias in quantifying judgments*. Hillsdale, NJ, USA: Lawrence Erlbaum.
- Schiffstein, H. J. N. (1995). Contextual shifts in hedonic judgments. *Journal of Sensory Studies*, 10, 381–392.
- Shaw, M. L. G. (1980). *On becoming a personal scientist: Interactive computer elicitation of personal models of the world*. London: Academic Press.
- Shaw, M. L. G., & Gaines, B. R. (1983). A computer aid to knowledge engineering. In *Proceedings of British Computer Society Conference on Expert Systems* (pp. 263–271) Cambridge, UK: British Computer Society.
- Shaw, M. L. G., & Gaines, B. R. (1987). KITTEN: Knowledge initiation & transfer tools for experts & novices. *International Journal of Man–Machine Studies*, 27, 251–280.
- Shaw, M. L. G., & Gaines, B. R. (1995). Comparing constructions through the web. In *Proceedings of CSCL95: Computer Support for Collaborative Learning* (pp. 300–307). Hillsdale, NJ, USA: Lawrence Erlbaum.
- Shaw, M. L. G., & Gaines, B. R. (1998, April). *WebGrid-II: Developing hierarchical knowledge structures from flat grids*. Paper presented at KAW '98: The Eleventh Workshop on Knowledge Acquisition, Modeling and Management, Banff, Alberta, Canada.
- Slater, P. (Ed.). (1976). *Dimensions of intrapersonal space* (Vol. 1). London: John Wiley.
- Steed, A., & McDonnell, J. (2003, Oct). *Experiences with repertory grid analysis for investigating effectiveness of virtual environments*. Paper presented at Presence 2004, Aalborg, Denmark.
- Suchman, L. (1987). *Plans and situated actions: The problem of human–machine communication*. New York: Cambridge University Press.
- Tan, F. B., & Hunter, M. G. (2002). The repertory grid technique: A method for the study of cognition in information systems. *MIS Quarterly*, 26, 39–57.
- Tomico, O., Karapanos, E., Levy, P., Mizutani, N., & Yamanaka, T. (2009). The repertory grid technique as a method for the study of cultural differences. *International Journal of Design*, 3(3), 55–63.
- Waterworth, J. A., & Fallman, D. (2003). The Reality Helmet: Transforming the experience of being-in-the-world. In *Proceedings of HCI 2003: Designing for Society* (Vol. 2, pp. 1–4), Bristol, UK: Research Press International.
- Waterworth, J. A., & Fallman, D. (2007, March). *Capturing users' experiences of interactive mobile technology*. Poster presentation at the British Psychological Society Annual Conference, York, UK.
- Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. Norwood, NJ, USA: Ablex Publishing Corporation.

Authors' Note

All correspondence should be addressed to
Daniel Fallman
Interactive Institute Umeå
c/o Umeå University, School of Architecture
SE-90187 Umeå, Sweden
daniel.fallman@tii.se

Human Technology: An Interdisciplinary Journal on Humans in ICT Environments
ISSN 1795-6889
www.humantechnology.jyu.fi