Dealing with User Experience and Affective Evaluation in HCI Design: A Repertory Grid Approach

Daniel Fallman[1], John Waterworth[2]

[1]Umeå Institute of Design
Umeå University
SE-90187 Umeå, Sweden

[2]Department of Informatics
Umeå University
SE-90187 Umeå, Sweden

ABSTRACT
In this paper we introduce, discuss, and empirically demonstrate ways in which the Repertory Grid Technique (RGT) may be used as a tool for dealing with user experience and affective evaluation. An empirical study using the technique in the context of mobile interaction is used to illustrate various uses and features of the technique and how it may be applied in HCI research and design.

Author Keywords
Repertory Grid Technique, User Experience, Affective Evaluation

INTRODUCTION

Recently, HCI researchers have begun exploring the relationship between user and artifact in terms of for instance its *affective qualities* rather than efficiency; *experiences* rather than performance; *fun* and *playability* rather than error rate; and *sociability and affective qualities* rather than learnability. What is yet undone, however, is find or develop appropriate and mature procedures for gathering and analyzing empirical data in relation to these new, experience and meaning-related aspects of interacting with computers. In this paper, we introduce, discuss, and empirically demonstrate a technique that could become a useful tool for dealing with the 'new HCI': *the Repertory Grid Technique* (RGT)

THE REPERTORY GRID TECHNIQUE

RGT is a technique for empirically eliciting and evaluating people's subjective experiences of interacting with technology. It may be regarded as a methodological extension of Kelly's *Personal Construct Theory* [17]. While it is not necessary to fully buy into the underlying theory to use RGT in practice, 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, which are bipolar in nature. According to Kelly, we judge for instance other people through forming construct such as *Tall—Short, Light—Heavy, Handsome—Ugly,* and so on. A 'construct' is hence a single dimension of meaning for a person allowing two phenomena to be seen as similar and thereby as different from a third [1]. Kelly suggested RGT as a structured procedure for eliciting a repertoire of these conceptual constructs and for investigating and exploring their structure and interrelations [1, 3, 18].

What is a Repertory Grid?

A *repertory grid* in itself is the outcome of a successful application of the technique. It is a table, a matrix, whose rows contain *constructs* and whose columns represent *elements*, i.e. the phenomena under investigation. Repertory grids also typically embody a rating system used to quantitatively relate each element in relation to the qualitative constructs. An individual repertory grid table is constructed for each subject participating in a RGT study.

First, an individual participating in an elicitation session produces his or her own constructs, i.e. what bipolar dimensions of meaning the person see as the most important ones for talking about the elements (the investigated phenomena). The construct elicitation process is typically eased by the use of *triads*, where the subject 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 their 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 own repertory grid table; which are quantitatively related to each other by use of some rating system. The constructs represent the qualities the participants use to describe the elements in their own, personal, words [10]; constructs thus embody the participant's *meaning* and *experience* in relation to the study's elements.

RGT in Human—Computer Interaction

Despite its popularity in fields such as organizational management, education, clinical psychology, and particularly in the development of knowledge-based systems [2, 20, 21, 22], the interest in it from an HCI perspective did peak in the 1980s, with a special issue devoted to the topic in the *International Journal of Man—Machine Studies* (Vol. 13, No. 1, 1980). But since then, the technique's appearance in HCI-related literature has been sparse, while not completely inexistent [e.g. 4, 15, 16].

AN EXAMPLE STUDY USING RGT: THE EXPERIENCE OF USING MOBILE INFORMATION TECHNOLOGY

In a study using RGT, we were interested in people's experiences and affective responses to mobile information technology, and to gain empirical insight into what kinds of meanings people ascribed to the different styles of interaction the various devices of the study embodied. To do this, 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.

In total 18 participants took part in the study, all of which had previously volunteered by signing up for a scheduled time slot. There were 14 (78%) males and 4 (22%) females. Eight of the participants were in the age span of 20—29 (44%), seven were 30—39 years of age (39%), while two were 40—49 (11%), and one was 50—59 (6%). On an preparatory questionnaire, three participants rated themselves as 3 on a 5-graded scale on self-estimated computer literacy (16%), 14 stated 4 (78%), while only one said 5 (6%). On a similar scale 1 to 5, when asked to rate their previous exposure to mobile information technology, one participant responded with a 2 (6%), 6 rated themselves as 3 (33%), 9 participant thought they were in the 4's (50%), while two considered themselves to be 5 out of 5 (11%). Each session lasted from 45 minutes up to two hours, averaging at slightly more than an hour. All participants took part in the study individually.

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. *The ABB Mobile Service Technician* (E5) is a wearable support tool for service technicians in vehicle manufacturing [5]. *The Dupliance* prototype (E4) is a physical/virtual communication device for pre-school aged children [8]. *The Slide Scroller* (E3) combines a PDA with an optical mouse to form a novel way of interacting with web pages on palmtop-size displays [7]. Finally, *the Reality Helmet* (E6) is a wearable interactive experience that seeks to alter its user's perceptual experience [6].

Construct Eliciting

Each participant sat at a table opposite to the experimenter. On the table, seven palm-sized cardboard cards where displayed. Every card contained a photograph of one of the four prototypes or one of the three existing 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—*triading* in RGT's technical language.

For each triad presented to her/him, the participant was asked to think of a property or quality that he or she considered notably enough to single out one of the three elements (devices) in the group, and to put a name or label on 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 then labeled E2's and E3's shared quality—as an opposite to 'Warm'—as 'Cold'. 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)*'.

The participant was then asked 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—also 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 seven-graded scale, where left (e.g. 1 of 7) would represented a high degree of the property found to be embodied by the singled out device (in the case of participant 10: 'Warm'), and right (e.g. 7 of 7) would represent a high degree of the property embodied by the two other devices in the specific triad (i.e. 'Cold').

Thus, for each triad exposed to a participant two kinds of data were being collected. First, a personal construct was elicited, i.e. a one-dimensional semantical space that the participant thought meaningful and important for discussing and differentiating between the elements of a triad. This process provides the study with *qualitative data*; insight into the participant's own meaning structures, values, and preferences.

Second, as each elicited personal bipolar construct was then used as *the scale* by which the participant rated all of the seven elements in the study, data was also gathered about the degree to which participants thought their construct had relevance to a specific element. This provided the study with *quantitative data*.

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 consists 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*.

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 likely have remained unnoticed if one had only looked for semantical similarity. This method may hence be called *semantically blind*, as it is driven primarily by each construct pair's quantitative data in relation to elements.

Second, what appear to be several semantically related and overlapping groups of construct pairs appear across the study's participants. Some similar bipolar scales, e.g. *'Young—Old'*, *'Appliance—Multifunctional'*, and *'Work—Leisure'*, can be spotted among several of the participants. It would hence be possible to go through the list of all participants' constructs and gather in groups those that bear semantical resemblance to each other, and analyze these groups (e.g. using discourse analysis). This approach could be regarded as *statistically blind*, as it is driven by an interpretation of the semantical content of the constructs, not taking the numerical ratings into account.

Participant Level Analysis

The manually collected data from the 18 participants was now compiled and inputted into the *WebGrid-III* application, a frequently used and feature-rich tool for collecting, storing, analyzing, and visually representing repertory grid data [12, 13, 14, 23]. 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* has been generated. Being 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* has been constructed for each participant. Here, both elements and constructs are sorted using the *FOCUS* algorithm [13, 14, 20] 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 vector space to facilitate maximum separation of elements in two dimensions [11, 24]. For more detailed information and discussion about these common ways of analyzing and visualizing repertory grid data, please see [13, 14, 20, 23].

Statistical Analysis of Multi-participant Data

For this study, we thought it would be interesting to see if there are any patterns or other kinds of relationships between different participants' repertory grids. To be able to perform statistical analysis on multi-participant data, all participants' 180 bipolar constructs were inputted 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 however immense and unstructured, so the task at this point became to refine and bring order into the fairly messy data set.

To discover such groups within the data set, the large, interpersonal repertory grid was subject to two cycles of FOCUS clustering. The difference between the two rounds is the manipulation of two rules that were being applied to distinguish groups or clusters in the data (for a detailed description of this process, please see [9]).

Naming Groups by Semantical Analysis

The 29 groups that resulted could be regarded as representing most pertinent dimensions of the participants' understandings of the elements of the study. The first task of this step was to create 29 new repertory grids based on the ingoing constructs of a group. The analysis, up to this point, had remained statistical rather than semantical. Thus, 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 was needed at this stage. 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 were chosen to be characterizing for the group as a whole, and used to form a new bipolar construct representing the group.

The hazard of possible experimenter biases and pure misunderstandings was reduced by choosing among existing labels rather than creating new ones to capture the character of a group. As an example of how this labeling was carried out, the group A16 consists of three ingoing 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.

Calculating Mean and Median Ratings

To be able to statistically analyze how the groups relate to each other and to the elements of the study, a rating for each construct on each element needs to be incorporated in the new repertory grid table. Rather than using arithmetic mean, these calculations relied on the median value. It has been found to provide a result which seems more true to the rating of the participants, where the influence of single, extreme values which are at odds with the majority of the values in the group become de-emphasized.

Interpreting and Presenting the Result

When applying an 85% threshold to these 23 clusters and their ratings (see [9] for details), the FOCUS algorithm further partitions them in 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 this additional clustering, 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 1) and as a PRINCOM map (Figure 2), which also shows how the different elements relate to each other. These ten dimensions are thus the most significant ways in which the participants experience the elements of the study.

DISCUSSION

While RGT is a theoretically grounded, structured, and empirical approach it is not restricted or limited to already existing, pre-prepared, or researcher-generated categories (i.e. constructs), as is the case with for instance Osgood's *Semantic Differential* [19]. In its first phase, RGT is clearly focused on eliciting constructs that are meaningful to the participant, not to the experimenter. In some sense, this means that the data which is found in a participant's own repertory grid has not already been influenced by—i.e. interpreted in the light of—the researcher's pet theory.

RGT is both Qualitative and Quantitative

As a repertory grid does not only consist of the personal constructs themselves but also of the rating of them in relation to the elements of the study, the researcher does not only gain insight into which the meaningful constructs are, but also—through the rating—the degree to which he or she thinks a particular construct applies or does not apply to a particular element. Hence, RGT technique that may perhaps be best characterized as being on the border of qualitative and quantitative research; a 'quali-quantitative' approach.

Results are Relational rather than Absolute

As that RGT relies on comparisons between different

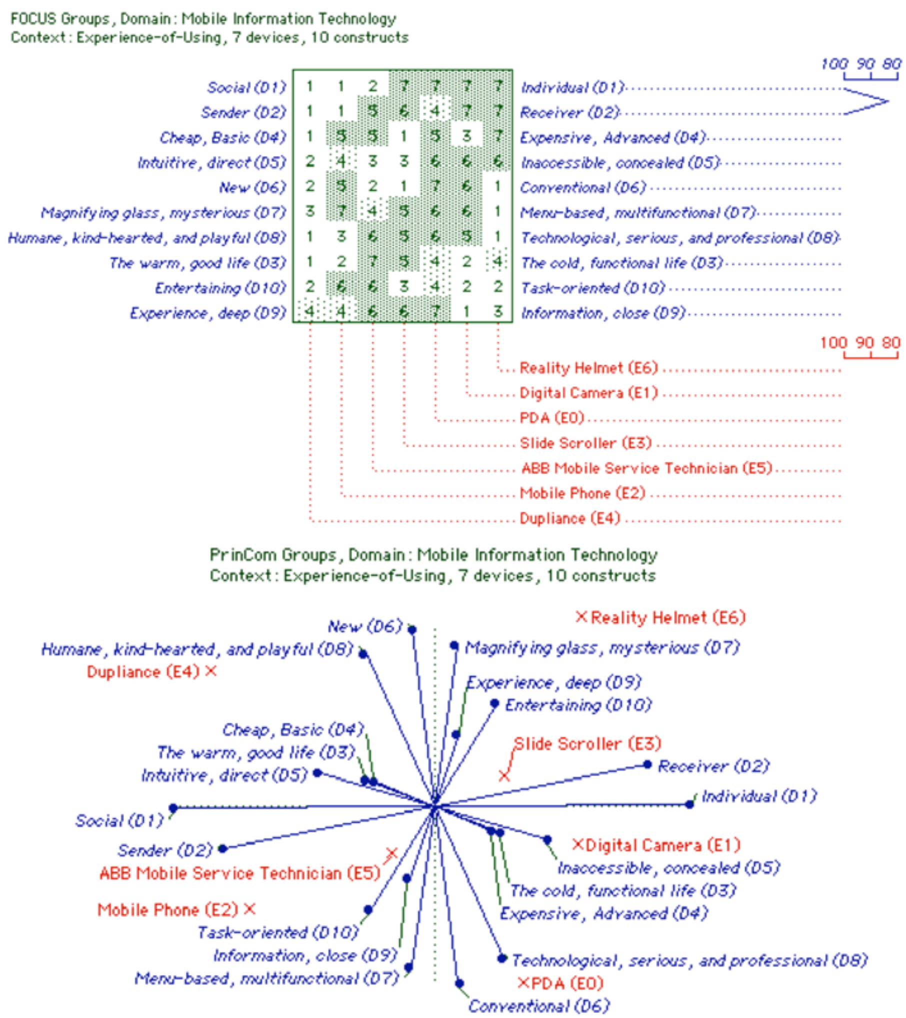

Figure 1 and 2. The ten resulting dimensions as a FOCUS graph (left) and as a PRINCOM map (right)

elements, all results must be regarded as relative to the group of elements that are included in the study. The result of a study using this technique is hence not a set of 'absolute values'. Rather, studies using RGT result in insight into people's experiences of things and the relationships between them.

Invested Effort

One expected disadvantage of RGT, at least when compared to fully quantitative lab experiments, is that it requires a substantial amount of effort to be invested both from the experimenter as well as from the participants at the time of construct eliciting. On the other hand, RGT is in some ways more efficient and less time-consuming than most other fully open approaches, e.g. unstructured interviews and explorative ethnography. As the personal constructs elicited from a participant 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 for instance unstructured interviews or ethnographic records.

CONCLUSIONS

The *Repertory Grid Technique* (RGT) has been introduced as an open and dynamic technique for eliciting qualitatively people's experiences and meanings in relation to technological artifact, while it at the same time embodies a possibility for that data to be subject to modern methods of statistical analysis. The RGT may as such best be described as a technique 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 a RGT study.

REFERENCES

1. Bannister, D. & Fransella, F. *Inquiring man.* 3rd Edition. Routledge, London, UK, 1985.
2. Boose, J.H. & Gaines, B.R. (Eds). *Knowledge acquisition tools for expert systems*. Academic Press, London, UK, 1988.
3. Dalton, P. & Dunnet, G. *A psychology for living: personal construct psychology for professionals and clients.* Wiley, London, UK, 1992.
4. Dillon, A. & McKnight, C. Towards a classification of text types: a repertory grid approach. *International Journal of Man—Machine Studies,* 33, 1990, 623—636
5. Fallman, D. Wear, point, and tilt: designing support for mobile service and maintenance in industrial settings, *Proc. DIS 2002,* ACM Press, 2002
6. Fallman, D., Jalkanen, K., Lorstad, H., Waterworth, J., & Westling, J. The reality helmet: a wearable interactive experience, *SIGGRAPH 2003,* ACM Press.
7. Fallman, D., Lund, A., & Wiberg, M. (forthcoming) ScrollPad: tangible scrolling with mobile devices, *Proc. HICSS'37,* IEEE, 2004)
8. Fallman, D., Andersson, N., & Johansson, L. come together, right now, over me: conceptual and tangible design of pleasurable dupliances for children, *Proc. CAHD 2001,* Asean Academic Press.
9. Fallman, D. In Romance with the Materials of Mobile Interaction: A Phenomenological Approach to the Design of Mobile Information Technology, Ph.D. Thesis, 2003, Umea University, Sweden.
10. Fransella, F. & Bannister, D. *A manual for repertory grid technique.* Academic Press, London, UK, 1977.
11. Gaines, B.R. & Shaw, M.L.G. New directions in the analysis and interactive elicitation of personal construct systems. *International Journal Man—Machine Studies,* 13, 1980, 81—116.
12. Gaines, B.R. & Shaw, M.L.G. Eliciting knowledge and transferring it effectively to a knowledge-based system. *IEEE Transactions on Knowledge and Data Engineering,* 5(1), 1993, 4—14.
13. Gaines, B.R. & Shaw, M.L.G. Concept mapping on the web. *Proc. Fourth International World Wide Web Conference*, 1995, O'Reilly.
14. Gaines, B.R. & Shaw, M.L.G. Knowledge acquisition, modeling and inference through world wide web. *Human Computer Studies*, 46, 1997, 729—759.
15. Grose, M, Forsythe, C., & Ratner, J. (Eds.) *Human factors and web development.* Lawrence Erlbaum Associates, Mahwah, NJ, 1998.
16. Hassenzahl, M. & Wessler, R. Capturing design space from a user perspective: the repertory grid technique revisited. *International Journal of Human—Computer Interaction,* 12(3&4), 2000, 441—459.
17. Kelly, G. *The psychology of personal constructs.* Vol 1 & 2. Routledge, London, UK, 1955.
18. Landfield, A. W. & Leitner, L. (Eds.) *Personal construct psychology: personality and psychotherapy*. Wiley, New York, NY, 1980.
19. Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. *The measurement of meaning.* Univ. of Illinois, IL, 1957.
20. Shaw, M.L.G. *On becoming a personal scientist: interactive computer elicitation of personal models of the world.* Academic Press, London, UK, 1980.
21. Shaw, M.L.G. & Gaines, B.R. A computer aid to knowledge engineering. *Proc. British Computer Society Conference on Expert Systems*, 1983, 263—271.
22. Shaw, M.L.G. & Gaines, B.R. KITTEN: Knowledge initiation & transfer tools for experts & novices. *Int. Journal of Man-Machine Studies* 27(3),1987,251—280.
23. Shaw, M.L.G. & Gaines, B.R. WebGrid-II: developing hierarchical knowledge structures from flat grids. *Proc. KAW'98,* 1998.
24. Slater, P. (Ed.) *Dimensions of intrapersonal space.* Vol 1. John Wiley, London, UK, 1976.