

Confessions from a Grounded Theory PhD: Experiences and Lessons Learnt

Dominic Furniss and Ann Blandford

University College London
UCL, Gower Street, London. WC1E 6BT. UK.
d.furniss@ucl.ac.uk; a.blandford@ucl.ac.uk
+44 (0) 207 679 7085; +44 (0) 207 679 0688

Paul Curzon

Queen Mary, University of London
QMUL, Mile End, London. E1 4NS. UK.
pc@dcs.qmul.ac.uk
+44 (0) 20 7882 5212

ABSTRACT

Grounded Theory (GT) is used within HCI research, but nuances and more modern interpretations of the method are rarely discussed. This paper has two intentions: to offer guidance on practical issues when applying GT, and to clarify the space of methodological possibilities. We describe an extended GT study on understanding why practitioners choose particular usability evaluation methods. We describe five stages in this study to highlight our experiences and choices made. We draw out seven practical and methodological considerations in applying GT in a CHI context. This challenges the more traditional inductive and objective positions on GT use; it sensitizes novices of GT to these issues; and through the extended case study it provides substance for debate on issues that affect those that use qualitative methods more broadly.

Author Keywords

Grounded theory, method, constructivist.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI):
Miscellaneous.

INTRODUCTION

Contemporary HCI is facing more qualitative research questions to engage with human values in a digital age [24], and develop methods, practices and designs ‘for the full range of human experience’ [30]. These challenges encourage better reflection on the art of performing different qualitative techniques to ensure strong research results are obtained. Grounded Theory (GT) is a method that can facilitate insight into people’s values, understanding and experience with technology. It has wide applicability to HCI; e.g. previous GT studies presented at CHI have included topics on immersion in video games [3], online trust and security [20], and password use [16]. Despite its popularity, nuances in the method’s use are rarely discussed in empirical papers, presumably because of

space limitations and an over-reliance on Strauss and Corbin’s [26] guidance on the method. These nuances are important: they have analytic consequences, and so should be recognised and considered. Furthermore, Matavire and Brown [18] call for more debate about GT use to reduce the widespread myths and misunderstandings around themes such as the role of literature and the use of prior theory.

At one level, this paper could be seen to confess a deviation from traditional approaches to GT. Some have argued that such deviations should drop the GT label and instead subscribe to doing qualitative research more broadly [27]. However, methods’ labels encompass their history, process and properties, which can be reflected on and critiqued. We offer guidance on practical issues and clarify the space of methodological possibilities around the GT method. These issues are pertinent to qualitative studies more broadly; e.g., should we just process data inductively, should we incorporate prior theory creatively, and how should this be managed? Arguably, these considerations make GT more relevant for contemporary HCI by going beyond a coding process to allow one to co-create understandings with users, and employ existing HCI theories to explore and elaborate findings. This paper’s format follows a reflexive account of GT use by Urquhart [28].

BACKGROUND

Grounded Theory (GT) was first proposed by Glaser and Strauss [10] and has since undergone developments and variations (discussed below). Broadly, it is a method with iterative data gathering and analysis. It uses coding processes heavily, but it is not just a way of coding. Importantly, it is different from methods such as protocol analysis and content analysis that emphasise reliability, validity and the counting of instances [22]. GT’s essence is in exploratory conceptual and theoretical development. It involves interplay between theoretical ideas and subjective understandings while requiring ‘fit’ with the data [23].

GT processes have been described in detail elsewhere [e.g. 1, 4, 13, 26]. Typically GT will involve interviewing a sample population about a topic, transcribing the interviews, coding parts of the transcript, and relating these codes to one another. Here, coding is a process of identifying and naming significant chunks of text. For example, I might label the previous sentence ‘coding’ and I

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$5.00.

might label this paragraph ‘GT method’. This form of abstraction facilitates analysis across different parts of the text or transcript and between different interviews. Analysis and coding should be carried out between interviews so that questions can be adapted to direct future data gathering. Variations include to what extent interviews and analysis are interleaved and the specifics of the coding process (e.g. not everyone will do open, axial and selective coding – this roughly translates to labelling codes, relating codes to each other, and selecting the main codes respectively [26]). Also, analysts with a more constructivist stance emphasise the use of analytic tools such as writing theoretical memos, defining codes and integrating codes into broader categories [23]. Here, a ‘constructivist’ stance emphasises creating a description of the context rather than discovering a description of it, which would be more ‘objectivist’.

Demonstrating the variation of GT use, a recent study in Information Systems (IS) identified four different uses of ‘grounded theory’ across 126 empirical grounded theory papers between 1985 and 2007 [18]. The four uses differ in their procedures and are dubbed ‘Glaserian’, ‘Straussian’, ‘mixed methodology’, and ‘technique application’. The Glaserian approach emphasises the emergence of ideas from the data away from established ideas and processes – a traditional inductive approach. The Straussian perspective introduces *a priori* coding procedures, frameworks and more directed questions [26]. The mixed methodology seeks to combine GT with other research methods, e.g. action research and case studies. Lastly, technique application does not show a strict adherence to any particular formulation but uses GT techniques, e.g. open coding, axial coding, selective coding, memos and diagramming. In terms of its output, GT does not have to build a total conceptual system with many concepts and connections; it can be used for: basic taxonomy development to show a hierarchy of classes and sub-classes in a context; focused conceptual development to describe a particular concept in detail; and cycles of interpretation between data gathering and analysis [23]. Henwood and Pidgeon [13] argue that it should not be presented as a unitary method.

The biggest controversy in GT variations is between Glaser and Strauss. Glaser disagreed with the development of procedural guidance and analytical frameworks, which he thought detracted from the original tenets of the method; i.e. these new developments add too much *a priori* shape, and inhibit the data’s ability to ‘speak for itself’ [12]. From this position, Glaser seems against the use of extant theory. Some authors claim that both Glaser and Strauss are against the use of extant theory [e.g. 18].

Charmaz [4, p. 134-5] criticises authors that reify statements in early works and turn them into static pronouncements, rather than treating them as starting points for an evolving method. She acknowledges that positions have not always been clear, but argues that the use of extant

concepts is allowed, although these concepts need to earn their right to feature in the analysis [4, p. 165-6]. Henwood and Pidgeon’s [13, p. 138] phrase ‘theoretical agnosticism’ is apt here. They say that “the special counsel that remains within grounded theory is to avoid being wedded to particular theoretical positions and key studies in the literature in ways that overly direct ways of looking and stymies the interactive process of engagement with the empirical world being studied.” We agree, and from the experience of our case study would say that the purposeful use of extant theory can be a source of creativity and insight, which a more traditional inductive approach would not afford.

Henwood and Pidgeon [13, p. 152] state that the challenge and excitement of all grounded theory research is getting out of the maze of detailed and complex codings, deciding on the limits to making constant comparisons, and reaching theoretical closure or integration. The case study here presents such a maze, focusing more on methodological issues than outcome, and on experiences and lessons learnt rather than results.

THE EXTENDED GT CASE STUDY

This case study’s research focus was on understanding why practitioners choose to use the usability evaluation methods (UEMs) they do. Urquhart [28] calls this sort of focus her ‘research problem’, which develops into sharper research questions as the study progresses. We were not aware of the benefits of a reflexive journal [22] during the study. However, other documentation is referred to, including weekly review meeting summaries, which helped us reconstruct this account. We highlight five stages in the study that were of significance for methodological reasons: before the beginning; acclimatising; fleshing out features; re-rendering the data; and theoretical lenses as tools.

Before the beginning

This study formed work towards the first author’s PhD in the UK, which is why it spans approximately three years. This is significant for understanding the study’s effort and duration. Like other studies, GT use is constrained by practical considerations, not least of which are deadlines for publication and when funding expires, which of course might raise tensions with purely methodological ideals such as theoretical saturation (see Lesson 1; all lessons are described in detail later in this paper). Theoretical saturation refers to a point in the analysis where one does not gain anything new from more data. When saturation has been reached, further data gathering is unfruitful.

We have already stated that the study’s research focus was on understanding why practitioners choose to use the UEMs they do; however, this was not the starting point. Months of reading, literature review, and discussion preceded this, which was performed to try to settle on a research problem that was suitable for a PhD and that fitted within the broad area of understanding design practice better. Competing topics included design expertise, design process, design

problems, and reflection in HCI. In the end, we preferred to try to understand why practitioners use some methods and not others. Essentially, academia develops lots of methods and few are adopted in practice, but the reasons for this were not well understood.

One approach to investigating this issue was to interview practitioners about why they chose some UEMs and not others. GT was already used by research colleagues, so this was recommended as an appropriate approach; the first author (hereafter referred to as A1) was advised to read Strauss and Corbin [26] to get started. It suited the project as an exploratory approach was needed, interviews were an obvious approach to the data gathering, there was guidance for the analysis, and others were already using it as a legitimate approach for HCI research.

One of the main alleged principles of GT is to start from the data and not prior texts, ideas and theories. With a mild concern that this had already been violated by doing an extensive literature review to settle on a topic (Lesson 5), A1 formulated a semi-structured interview script which covered five themes:

- *Background* of the participant.
- *Work organization*, concerning how the work is organized, the structure of the organization, whether there are teams, project lifecycles and challenges.
- *Business client relationship*, on communicating with clients for attracting, doing and handing off work.
- *Practitioner skills*, concerning what skills are needed to be a good practitioner.
- *Tools and techniques*, concerning what methods are used, how they are used, when they are used and what is valued in a method.

Without the extensive literature review, prior education in HCI and the supervisor's experience, we would not have been able to start with a developed interview script or settle on a topic of such theoretical interest (Lesson 5).

Acclimatising

The deadline of a first year report viva focused attention on doing some preliminary data gathering and analysis within the first 9 months. This pragmatic driver also had methodological significance by providing a stage of review.

We chose HCI practitioners in the website domain as a sample population to focus the study. The first four participants were known to the researcher, were easily available, and allowed A1 to ease into the interviewing and subject matter in a more comfortable way. We also thought that we would be able to get better data from more experienced practitioners later in the study after our understanding had matured. The first four interviews had between 1 and 3 years experience in industry. This sampling therefore did not conform to the idealized

processes of theoretical sampling but was influenced by pragmatic constraints (see Lesson 1). Theoretical sampling is a sampling technique that is based on what is best for developing the theory, e.g. perhaps by selecting contrasting participants or someone that could elaborate a particular feature of the context. It is very considered so is far from random, and in its purest form neglects pragmatic influence.

These early semi-structured interviews provided a learning opportunity, enabling A1 to get used to the variations between interviewees' styles, in handling questions and probing. At the start of the interviews, A1 took great pains to ask open-ended questions and not put words in the mouths of interviewees, but it became apparent that this population was independent enough not to warrant such concern, and it appeared that a relaxed, discursive interview style opened up the interview to yield better data. It was also apparent with successive interviews that A1 brought forward the points, words and examples of the previous interviews, building on each conversation (see Lesson 2).

Following the advice of colleagues, A1 proceeded with open coding [26] and used Atlas.ti software to support the process. It was difficult to know at what level to code the data, but through the process it became apparent that the level initially chosen was too detailed, as the list of codes soon became too long and unmanageable (Lesson 3).

This early analysis developed a narrative that centered around three main concepts: power, persuasion and value added HCI. These were extracted by trying to get a sense of the main themes thus far. Each theme was supported by quotations as the extract from A1's first year viva report shows (Excerpt 1; participant quotations are in italics).

Fleshing out features

We fed the 4 interviews from the preliminary study into the main study phase, interviewing 9 HCI practitioners from the website domain in total. To explore the potential differences in culture we interviewed six practitioners with experience in full service agencies, three in usability consultancies and three with in-house service experience. Some had experience in multiple areas. Total experience for the three contexts was over 15 years, over 10 years, and 3 years respectively. We interleaved data gathering, coding

Excerpt 1. Description of 'Power': "Power is an important concept as there is always someone to report to, someone that is managing your work and someone that is paying for your work. This power normally resides in a hierarchy with it being concentrated toward the top; clients are likely to hold the power in a consultancy relationship because they are paying for the service, and management and senior staff are likely to hold the power when in-house.

R4: ...at the end of the day it is the client who is the one that is paying the cheques so it's up to them what they want really..."

and analysis where practical, but were constrained by when it was convenient for participants to meet. Some interviews were a week apart, some a few months; 8 interviews were completed in seven months and the last much later.

The interviews kept building on each other, and the open discursive style sometimes led to apparent contradictions and clarifications that allayed fears of inadvertently biasing the interviews; for example participant R8 corrected an implicit assumption made by A1:

“I: I wondered what sort of things clients come to you for, [...] an obvious one might be to increase revenue [...] if it were an ecommerce site, but I wondered whether there were any more reasons that come to mind?”

R8: Yeah, it would range... well the unspoken assumption behind that question is that all the clients know why they have come to us, and they don't, sometimes the biggest portion of our job is to work with them to figure that out.”

Rather than tidy the coding scheme of the previous stage, we recoded the data anew at a more manageable level. Moving on to the other interviews, we attempted to perform the coding stages consistently with Strauss and Corbin [14], and in keeping with the Glaserian style of inhibiting and excluding extant theory. We focused on developing the GT inductively. With hindsight, coding the first 4 transcripts appears to conform to Charmaz's [4] 'initial coding' stage, where transcripts are analysed in an unfocused manner. As subsequent interviews matured, we moved on to more 'focused coding', elaborating concepts and themes. However, it was hard to perform selective coding as several categories were competing for dominance. In the end, four categories provided the corner stones of the narrative [9]: methods and processes; relationships; communication and coordination; and psychology and expertise. In terms of presentation: after a closer introduction to the analysis and supporting quotations, the four dominant categories were described and linked to the existing literature (see Excerpt 2 for an example).

Re-rendering the data

The four categories that emerged from the previous stage

Excerpt 2. Description of 'Methods and Processes': "The analysis has shown that usability work is heavily influenced by the clients' needs. This commercial focus puts emphasis on effective and pragmatic choices that will deliver results to agreed time and budget scales. This is reflected in Wixon and Wilson's [29] move away from science to "the art of the possible under constrained resources" in usability practice; and Cockton's [5] claim that HCI should be more about delivering value than finding the truth. This is perhaps what one participant meant when distinguishing scientific validity from commercial and design validity. [...] [9, p. 162]

(above) provided a rendition of the data, but it left the analyst wanting. From the nine interviews this rendition highlighted areas of importance. However, it did not show the diversity between interviews as it was an aggregation, it did not help readers to become intimate with the data, and while the hierarchical format of the four high-level categories highlighted areas of attention it felt disjointed. The question remained as to whether there was a way of getting readers closer to the data, demonstrating the diversity between interviews and integrating themes better, in a narrative, so that they could be more easily understood and communicated. With hindsight, these concerns of how to best render the data were the first signs of moving from 'how do I represent the respondents' views validly?' to 'how do I best render the data to communicate a useful message to an audience?' (see Lesson 6 on objectivist versus constructivist GT). Taking this forward this next phase intentionally experimented with alternative treatments of the data to explore options, strengths and weaknesses in how to render it.

Following the principle of theoretical sampling, we decided to expand our data set to include human factors (HF) practitioners in the safety-critical domain. We interviewed 13 HF practitioners, from 8 different work places. Six practitioners had between 1 and 5 years of HF experience and six had between 6 and 30 years of experience. One engineer had no direct HF work experience but described managing a project that had a strong HF component. A1 started by interviewing practitioners that were known and more accessible to the analyst. This domain was less familiar to the analyst. A literature review was carried out to develop a better vocabulary and to make A1 familiar with the issues, methods and practices that participants might discuss (Lesson 5). Again, data gathering and analysis were interleaved where possible. On one occasion the analyst travelled and interviewed three practitioners from the same company in one day, which made practical sense (Lesson 1).

Three treatments of the data were trialled. The first was a summary of each interview. Two of the summaries can be seen in Excerpt 3. These gave an idea of the breadth of data but lacked insight that generalised across cases.

The second treatment was a view of the coding network. Atlas.ti allows the analyst to link codes together in network diagrams to explore relationships between codes. The intention here was to give the reader access to the codes and links that had emerged from the analysis. However, the full GT network was overpopulated (128 codes) and too dense to readily make sense of. A simpler, higher-level network was created by using 16 codes that were significant due to the number of quotations associated with them, the number of links to other codes and the number of participants that mentioned them. Figure 1 shows this diagram merely to

Excerpt 3. Variance in participant data shown through interview summaries of S1 and S2:

Respondent S1: Here design solutions were driven through iterations with input from people with knowledge of the products and working practices, rather than the specific identification of safety issues through evaluative methods. Much of the communication is captured in design drawings and so documentation is in pictures and notes rather than wordy reports. Even though they work in-house they still have to sell their ideas and services, and face the same issues of not being involved or being involved too late that out-house people face. The design-solution focus forces them to engage with the real trade-offs. They apply patterns through analogical reasoning to aid the design process, i.e. they are familiar with reoccurring issues that inform designs.

Respondent S2: This contrasted with solution focused consultancies in that it was quite formal, independent and research driven. Rather than taking a design orientation the work appeared to be very evaluative, a lot of it taking the form of controlled experiments where safety could be independently evaluated. Reports were written in a similar way to research reports that you might find in academia. Written communication seemed to dominate client contact so an audit trail was maintained and misunderstandings reduced. The rigor of their research and independent status characterise the company's offering. Often they do not know what happens to their results and subsequent designs as they are detached from the process. Expert panels and discussion groups were recognised as useful methods for tapping into domain expertise. [7, p. 119]

illustrate what such a network looks like. Each of the codes in the diagram was described with reference to the other codes it was linked to (see Excerpt 4 for an example). However, the presentation of this network and descriptions, although allowing one to become more intimate with the analysis, lacked a core focused message for the reader to take away.

The third data treatment involved developing a narrative around the concept of 'downstream utility'. Traditionally, researchers were focused on the implementation of the UEM, but more recently attention has turned to how information from UEMs can better transfer value 'downstream' [e.g. 17]. Our data also showed that UEM choice and use depended on factors 'upstream', such as resource allocation for the projects and past experiences between the client and consultant. A river metaphor started to develop and another creative jump was made, using what with hindsight would be called analogical or metaphorical reasoning. The stream or river was in a landscape of context-shaping factors composed of social factors (e.g. client-consultant relationship), technical factors (e.g. specific HCI issues), the project's structural factors

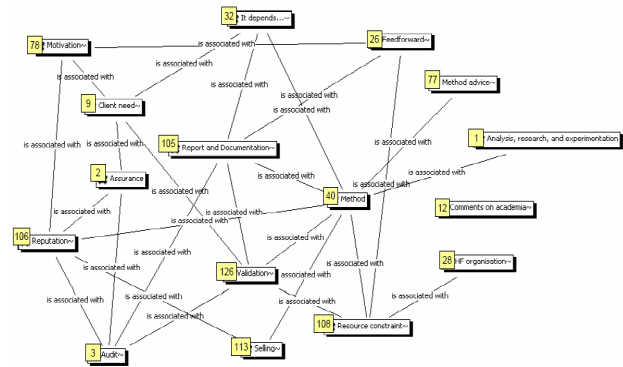


Figure 1: Network diagram of 16 codes and links [7, p. 125]

Excerpt 4. Description of the code 'Reputation': There can be the reputation of HF in general, the HF organization, the HF practitioner, methods and ideas; and this can be influential in organizational decision making.

Reputation has to be worked for and quality maintained. The reputation of a practitioner will facilitate selling their services as it will provide the client with some reassurance that the work will be completed to a good standard and their recommendations will be sound. The expertise of the practitioner will be linked to their reputation, and greater experience will reduce the risk of a project failing. New practitioners, new methods and new practices that have a weak track record will make a project less predictable. There will be a motivation and prejudice to select practitioners and methods that they have confidence in.

Practitioners and organizations can be audited by clients and regulators to check their quality, which will influence their reputation. Good work will more likely lead to repeat business and attract more work. [7, p. 372]

(e.g. the design stage), communication factors (e.g. reporting back to developers or the chief executive) and resource factors (e.g. the capabilities of the practitioners and the time for the project). These context-shaping factors were derived from the data. The metaphor led to the development of a 'conceptual picture' (Figure 2) that aided the sense-making of the analyst and appeared to provide the more coherent narrative for readers that he was looking for. For example, this appeared to provide a more fluid story than the presentation of the four categories introduced in the 'fleshing out' stage (above). The use of the metaphor led to a focus on 'process' rather than only 'themes' as recommended by Charmaz [4]. Now we have a *story* around events rather than simply a hierarchical list of related factors.

Theoretical lenses as tools

Analysis of a more top-down nature, e.g. through the use of pre-existing literature and theory was inhibited in the previous three stages where we emphasized an inductive or bottom-up approach (see Lesson 5). In the theoretical lenses

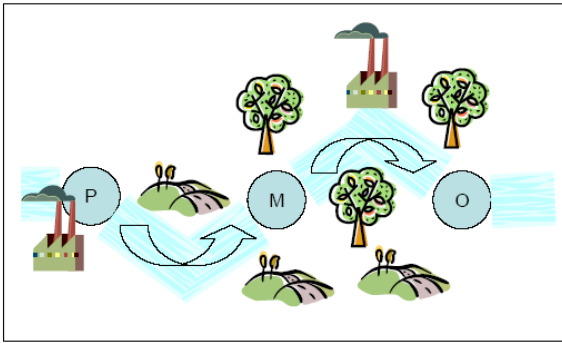


Figure 2: To show that project planning 'P', method application 'M', and project output 'O' are linked together by a stream that is shaped by context-shaping factors

stage we stopped gathering empirical data and explored the leverage that extant theory could provide in interpreting our data. It involved two separate stages of theory bridging, the first using Distributed Cognition (DC) [14] and the second using Resilience Engineering (RE) [15].

Preparation and selection of extant theory

In terms of identifying a potentially useful theoretical framework to apply, these two stages involved serendipity and a 'prepared mind' [6]. For example, DC was a body of work that the analyst was familiar with because of previous projects. This meant the analyst's mind was prepared to see bridging opportunities between emerging concepts from the data and DC theory. This acknowledges that researchers will inevitably come to a situation with some established ideas and preferred theories [19], and that the researcher is integral to the interpretation of the data [4]. Serendipity played a large role in highlighting RE as a fruitful extant theory for bridging purposes. RE was only encountered when preparing to talk to HF practitioners in the safety-critical domain, and by having a colleague highlight RE as a related area. After the re-rendering stage of analysis, the analyst was intimate with the data and more prepared to see bridging opportunities when extant theories and concepts were encountered.

Creating links between 'theory and data' / 'data and theory'

A1 took a systematic approach to explore the creative links from data to theory and back. He externalised this process sufficiently so that it could be developed. Reviewing the literature of the extant theory was an important step to 'prepare the mind' to recognise links between the theory and the data. Potential links between the extant themes and the emerging concepts were externalised through notes and memos [4; 19], which are essentially written notes of thoughts about the developing GT analysis (Lesson 4). The constant comparison method [22] then led to these links being explored with further codes, categories, descriptions and information from the literature. As we progressed with this approach we adopted a template for each time we made a bridging step between the data and the extant theory: we

first explained the theme in the theory, we then explained how this was reflected in the data, and finally we discussed broader observations of this link (e.g. see Excerpt 5). One must also be open to apparent contradictions between theory and data, which will highlight limitations in linkages. Trying to understand discrepancies can lead to further investigation and insight.

DC [14] concerns itself with the propagation and transformation of information in systems. Consequently it provided leverage by embellishing the factors influencing information flows of HCI/HF project work; e.g., transformation of information could be tracked from the client need, to HCI/HF project aims, to user testing design, to user testing video recordings, transcriptions, analysis, project reports and distilled recommendations. Furthermore, this extant theory provided an opportunity for more mature reflections on developed concepts such as boundary objects (Star, 1989), organisational expertise and epistemic actions. For example, boundary objects were recognised such as the client's need that was reinterpreted in HCI/HF terms, the contract between the client and the HCI/HF practitioner, and the final report. Some practitioners wrote the latter with the audience in mind, i.e. knowing that it may have to satisfy the different needs of regulators, directors and developers.

Resilience relates to safety, but focuses more on unexpected events and how systems cope, adapt and recover from these, e.g. rather than a reduction in error per se. RE [15] resonated well with the data from the bottom-up stages. For example, RE favours systemic over linear explanations of accidents, has adaptation under constraints as a focus, is interested in normal work as well as extreme situations, and emphasises understanding phenomena from the perspective of those that are inside the context. These elements provided the initial hook for the analyst, and enough interest to explore bridging data with theory in more detail (e.g. see Excerpt 5). Interestingly, there was some back propagation to RE theory suggesting a different emphasis to its normal use; i.e. the avoidance of accidents. This was a modification of the extant theory which was inspired by our data and which came to the fore by the constant comparison method [22]. For example, the non-linear explanations of accidents resonated strongly with the data but on closer reflection the theory did not entirely match. This did not mean it was inapplicable; instead it led to a moment of creative tension to work out what this meant for the data and what this meant for the theory. Here, RE would normally focus on how events negatively interact to cause accidents and look to dampen these negative interactions to stop things going wrong; we instead found it more useful to consider how HCI and HF practitioners maximise positive interactions in a project so that things go right with limited resources. This led to a switch of focus from 'negative resonance' to 'positive resonance' [7, 8].

Excerpt 5. Bridge to concept of ‘Loose Coupling’:

Theory: Grote [11, p. 116] states that “a core requirement for resilience is to achieve an adequate balance between stability and flexibility in the functioning of an organization.”

Support: This is evident in the labeling of techniques and methods that add stability to a design project, and where their practice can be adapted to suit the context. For example, Heuristic Evaluations [21] were reported to be used in an ad hoc manner to support design recommendations, explicitly used to evaluate and compare websites, implicitly used like an expert evaluation, and actual heuristics were sometimes adapted from “Nielsen’s ten heuristics.”

Discussion: The loose coupling evident in labeling simplifies communication of project elements and structure to clients. According to our interviewees, novices (e.g. clients) are less able to cope with the details of potential project variances. Labels and prescriptions help overcome this. [8]

Further analyses using extant methods

After these types of theoretical links between themes were made, a further stage of rendering the data was explored by using analysis methods from the respective theories. For DC this led to the use of DiCoT (Distributed Cognition for Teamwork [2]) and for RE this led to the use of FRAM (Functional Resonance Accident Model [15]) – both methods that support the analysis of complex systems. DiCoT is a method normally used to analyse observational data of teamwork settings in terms of DC. This method involves developing five models of the system: the social, information flow, artefact, evolutionary and physical models [2]. This can involve diagramming and reference to DC principles. In this case, the interview data provided the information to populate the models rather than observational work. FRAM identifies and maps functional elements in the system and how they influence each other [15]. Both of these methods developed the role of ‘process’ at the core of the GT, and lent their theoretical weight to exploring the nature of the complex system in which this process was situated. Again, preparing the mind to recognise links through familiarity with the data and the method, extensive use of notes and memos, and the constant comparison method to check the legitimacy of insights created the conditions necessary for this analysis.

These methods provided an analytic framework and theoretical insight that extended the original analysis in the more bottom-up stages. Using the FRAM analysis as an example, six subsystems were identified as influencing the choice and use of usability evaluation methods: the project process; HCI/HF practitioners understanding; persuasion, rapport and reputation; staff development and management; tools, methods and reporting practices; and auditing and documentation. These overlapping subsystems involved 29

individual functional elements, and their representation resembled a GT network diagram. The picture that this theory supported, which had strong resonance with the data, is that there are many factors in the context that can influence the adoption and adaptation of method use in complex ways. Some of these factors can be at the ‘sharp-end’ [15] close to the point of method use, e.g. a specific request by the client or the budget they have; and others can be at the ‘blunt-end’, far away from the point of method use, e.g. what practitioners were taught and industry trends.

Validation through member checking

To check that the 29 functional elements and 6 subsystems we had constructed were recognisable by practitioners, 10 of the 22 participants gave their feedback on them, as well as 8 practitioners not involved in the interviews [7]. This showed broad support for the work with some suggested modifications (see Lesson 7).

LESSONS LEARNT

Seven major lessons emerge from the above case study:

Lesson 1. GT studies are managed alongside their practical constraints, and so should not be appreciated against imagined objective ideals. Consideration needs to be given to the scope of the study and its output, the numbers of participants and their accessibility, and the depth of the analysis. One way this came in to play in our GT was the length of the study, which extended over a three-year period, but was also punctuated by milestones (vivas) at which interim findings needed to be reported. Other projects may be much smaller and so not have the time to experiment with data treatments and develop a conceptual framework. This extended duration allowed time for the researcher to mature with the method’s use, which is important when GT is so entwined with the analyst using it.

Practical constraints also came to the fore when organising interviews. Theoretical sampling suggests that you select future interviews and data points to test your emerging theory, so exploring the boundaries and diversity of concepts is key. This was practiced in selecting HCI practitioners in the website domain with different experience, and from different companies, e.g. in-house, full service agencies and independent consultancies. The more diverse comparison was between the website and safety-critical domains. However, practitioners that were more accessible were interviewed more readily, which follows a convenience sampling model rather than a theoretical sampling one.

Theoretical saturation occurs when new data is not leading to new insight or further conceptual development [13]. This is associated with the last stages of a GT, when the analysis and picture are complete. However, when this actually happens appears to be subjective, a decision which could be encouraged by a looming deadline. Also, ‘saturation’ can be problematic from a constructivist perspective because the data can be reinterpreted in new ways [e.g. 4].

Lesson 2. Interview style must be adapted to the situation but open and friendly is best. Initially, there was an acute concern over asking open-ended questions that would not bias the interview, but in hindsight this was misplaced. For our population, participants were professional, articulate, would defend their views, and made their experience understood. As the study developed, the analyst adopted a more discursive style of interviewing, trying to create a friendly and respectful atmosphere in the interview, to open it up and get good data. This became a lot less like a question and answer session and more like a discussion between equals where a joint understanding was created. This became easier to do further into the analysis where the analyst had the foundation of previous interviews' examples, experience and opinion. It was felt that creating a relaxed atmosphere, e.g. through trust and humour, worked well in opening interviews up; but of course not all interviews were as open as others due to confidentiality issues, familiarity with the analyst and just the way different personalities interact.

Lesson 3. Start coding low and move upwards. Pause early in the process and reflect on the coding level. Through experience we found that the granularity of coding should adapt according to the centrality of the text to the emerging themes, from leaving paragraphs of transcript untouched where it was not relevant, to going through particularly interesting sentences word by word. Adams et al. [1] and Urqhart [28] show examples of GT coding practice.

We would advise novices to start too low and come up as we did, rather than too high and come down, as otherwise too much detail can be glossed over. However, in the end the analyst has to feel their way through the codes. We paused after the first 4 interviews, which gave time to reflect on the analysis and the granularity on the coding scheme. This early pause was useful and will likely be useful for future studies. Charmaz [4] calls this the 'initial coding' stage, where transcripts are analysed in an unfocused manner. After this stage, the coding scheme either has to be tidied up or started again with the benefit of being acclimatized by the preliminary analysis. We started again, and as subsequent interviews matured we moved on to more 'focused coding', fleshing out concepts and themes.

Lesson 4. Use analytic tools and use them flexibly. At the start of this study, the prevailing view was that GT analysis centred around the coding mechanisms of Strauss and Corbin [26], i.e. open coding, axial coding and selective coding. The mechanisms and benefits of other analytic tools, such as memos, and network diagrams, were neglected. The role of these tools, supported by the Atlas.ti software, grew as the study developed. This in part came from a confidence that GT appeared to be less about selecting and applying a relevant coding scheme so that it accurately represented the data, and more about making sense of it, linking it to different ideas and thoughts that occurred in the analysis and creating an insightful rendering

of the data that would be understood by and useful to others. The bigger conceptual moves in our analysis, involving the development of the stream and landscape metaphor, and the links to DC and RE, were explored through memos and network diagrams. Analytic tools should be used flexibly as long as it helps the analyst understand the data, develop ideas, and create insights that 'fit' the data. Just relying on codes leaves the analyst somewhat limited to hierarchies of codes, categories and themes that may fall short of a conceptual framework or miss those creative insights that could open up an area and move its understanding forward.

Lesson 5. A top-down constructivist approach should not be neglected. It has potential for better questioning, theory integration, insight and a richer picture compared to a more traditional inductive approach. From the outset of the study, the prevailing view was that GT was a strict inductive method to create theory about some context or phenomena from participants' perspectives. However, a purely inductive view is problematic as no analyst is a *tabula rasa*. Through our analysis we moved from a position of trying to inhibit the use of extant literature and theory, to intentionally using it for analytic and creative purposes.

Before interviewing practitioners in the two different domains we performed an extensive literature review. This was necessary to select the research problem of the study, to develop a reasonable semi-structured interview script, to internalize their vocabulary and to become familiar with issues that the participants would discuss (particularly in the safety-critical domain which was new to A1). This was necessary to make the interviews engaging and productive; less preparation would have jeopardized this.

The top-down approach of the theoretical lenses stage could be seen to be controversial when considering the traditional advice to inhibit and exclude extant theory. However, in our analysis we found that the bottom-up analysis benefited from existing theoretical concepts and structures. This is less controversial for those that argue for a more modern constructivist revision of GT [e.g. 4]. Here, the same principles that should be applied in any GT analysis, e.g. emergence, constant comparative method, and theoretical sampling [18], continue to be important. The interpretation must remain faithful to the data, and favour emergence over forcing [e.g. 4 & 13].

In the experience of our study the top-down phase strengths include: getting vertical validity by referring to abstract concepts that have been noted elsewhere; using extant theory as leverage for insight and creativity that can bring new perspectives on different contexts; breaking writer's/analyst's block; and developing extant theory by resolving discrepancies between this and the data. Potential weaknesses include: warping an objective view of the world; allowing the researcher to bias and influence the data and analysis; and inhibiting what the data says about itself. Overall, our case study shows an extended inductive

process that has benefited from using theoretical lenses as tools to explore the data. These top-down mechanisms can bring benefit if managed appropriately, by helping to construct a better rendering of the data. Constructivists will more readily accept the use of extant theory for helping to see the data in a new way; whereas more objectivist views will emphasise seeing the data on its own, as a window on an objective world that should not be biased by the analyst or extant theory. In our experience, constructivist revisions to GT can yield a richer and more insightful analysis.

Lesson 6. Appreciate the multiple styles and purposes of GT. Henwood and Pidgeon [13] stress that the excitement and challenge of GT is finding a way out of its maze, but there is no one legitimate way out of the maze. GT is associated with building theory, and at the most developed level this includes a full conceptual system, but at lower levels of development will also include: basic taxonomy development, focused conceptual development and cycles of interpretation [23].

All stages of our study included cycles of interpretation. The fleshing out stage involved thematic analysis. The re-rendering stage involved the development of a conceptual system through the stream and landscape metaphor. The theoretical lenses stage involved conceptual development, integration with extant theory, and the creation of a conceptual system. So choices and trade-offs have to be made in terms of what is done in a GT and what form of output one will have when out of the maze. It is scientifically responsible to remain theoretically sensitive [4; 13] so the GT relates to, builds on and challenges previous work. Neglecting this risks ‘reinventing the wheel’ or wondering how a GT fits the literature after its analysis has been complete. This is where the importance of the study’s findings come to life or we are left asking, ‘So what?’

Our stance on the purpose of GT changed with our increasing resonance with a more constructivist revision, moving away from more objectivist views. From our understanding the purpose of a GT study can change dramatically depending on whether an objectivist or constructivist stance is adopted, i.e. the former seeks to represent each participant’s views accurately and validly like a privileged window into the participant’s world, and the latter tries to create a rendering of the data that communicates an insightful, valid and useful message to an audience. This difference impacts on the way one conducts a GT, what one is left with when out of the maze, and the criteria for assessment.

Lesson 7. Assessment should fit the style of the project. GT fruit is often produced by an individual, following a style and within a scope, which we have detailed in our case study. However, it could be by a team, applying an emergent coding scheme, or over a shorter period of time. Pidgeon and Henwood [23, p. 101] have argued that the stance and purpose of a GT will impact on the criteria to

assess it. However, they offer four questions to orientate an assessment:

1. Was respondent validation attempted?
2. How well documented was the process of analysis?
3. Were concepts linked together in justified ways?
4. Was the problem of reflexivity addressed?

From a constructivist perspective respondent validation or member checking seems the most developed form of assessment because this enables participants (subject-matter experts) to check the final rendering of the data. Checking the output is especially important where some of the analytic process may be hidden. For example, the process of arriving at a rendering might be a complex and creative process involving theory, data and ideas whose integrated mechanics might be indescribable and so uninspectable. For our study HCI/HF practitioners checked different outputs at three stages including the RE analysis in the theoretical lenses stage, but they did not check all the parts because of the time and effort it would take them. From a constructivist perspective, inter-rater reliability is misplaced as a method of assessment because GT is not the application of a coding scheme, as a contents analysis would be; instead it is the exploration and rendering of data in a way that provides conceptual insight into some context or phenomenon. Writing up a GT, reporting results, and sharing enough of the data to satisfy readers who may be from different methodological stances remains a challenge and one that we are still learning.

CONCLUSION

GT is a flexible qualitative method that can be used for basic taxonomy and conceptual development, and the creation of theoretical frameworks. More modern constructivist revisions of GT [e.g. 4] move away from traditional data-driven approaches, which seek to capture an objective view of the world. Instead they offer flexibility to co-create understandings with users and can employ HCI theories to explore and elaborate findings in a more top-down fashion (e.g. by using DC and RE). In our experience extant theory provided theoretical weight, insight and was a source of creativity. This paper provides guidance on practical GT issues through the case study and lessons learnt; it has also clarified the space of methodological possibilities, which has consequences for style, purpose and assessment of the research. This will hopefully inspire and guide those choosing to embark on related studies.

ACKNOWLEDGMENTS

We are grateful to all participants who contributed their time and expertise to this study; to Helen Sharp and Alistair Sutcliffe who provided feedback on an earlier version of the paper; and for the anonymous reviewers’ contributions. This research was funded by EPSRC Grants GR/S67494/01, GR/S67500/01 and EP/G059063/1.

REFERENCES

1. Adams, A., Lunt, P. & Cairns, P. A qualitative approach to HCI research. In P. Cairns & A. Cox (Eds.) *Research Methods for Human-Computer Interaction*. Cambridge University Press. (2008).
2. Blandford, A. & Furniss, D. (2006). DiCoT: a methodology for applying Distributed Cognition to the design of team working systems. *Proc. DSVIS 2005*, LNCS, (2005), 26-38.
3. Brown, E. & Cairns, P. A Grounded Investigation of Immersion. (2004). *Proc. CHI 2004*.
4. Charmaz, K. *Constructing grounded theory*. Sage Publications, London. (2006).
5. Cockton, G. (2004). Value-Centred HCI. In *Proc. NordiCHI 2004*.
6. Foster, A. & Ford, N. Serendipity and information seeking: an empirical study. *Journal of Documentation*, 59. (2003). 321-340.
7. Furniss, D. Beyond Problem Identification: Valuing methods in a 'system of usability practice'. PhD Thesis. UCL. (2008).
8. Furniss, D., Blandford, A. & Curzon, P. Resilience in Usability Consultancy Practice: the case for a positive resonance model. *Proc. Resilience Engineering Workshop, Linköping Electronic Conference Proceedings*. Sweden, 23, 5, (2007) 31-35.
9. Furniss, D., Blandford, A. & Curzon, P. Usability work in professional website design. In E. Law, E. Hvannberg, & G. Cockton (Eds.) *Maturing Usability: Quality in Software, Interaction and Value*. (2008).
10. Glaser, B. & Strauss, A. *Discovery of grounded theory*. Sociological Press. (1967).
11. Grote, G. Rules management as source for loose coupling in high-risk systems. In *Proc. of the Second Resilience Engineering Symposium*. (2006). 116-124.
12. Heath, H., & Cowley, S. Developing a grounded theory approach: a comparison of Glaser and Strauss. *Int Journal of Nursing Studies*, 41, 2, (2004), 141-150.
13. Henwood, K. & Pidgeon, N. Grounded theory in psychological research. In P. Camic, J. Rhodes, & L. Yardley (Eds.), *Qualitative research in psychology: Expanding perspectives in methodology and design*. APA, Washington DC. (2003).
14. Hollan, J., Hutchins, E., & Kirsh, D. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.*, 7, 2, (2000), 174-196.
15. Hollnagel, E. *Barriers and Accident Prevention*. Ashgate Publishing, Ltd. (2004).
16. Inglesant, P. & Sasse, A. The True Cost of Unusable Password Policies: Password Use in the Wild. *Proc CHI 2010*.
17. Law, E. Evaluating the Downstream Utility of User Tests and Examining the Developer Effect: A Case Study. In *IJHCI*, 21 (2). (2006).
18. Matavire, R. & Brown, I. Investigating the Use of "Grounded Theory" in Information Systems Research. *Proc. SAICSIT 2008*. ACM Press (2008).
19. Miles, M., & Huberman, A. *Qualitative data analysis*. Sage Publications. (1994).
20. Nilsson, M. Adams, A. & Herd, S. Building security and trust online. *Proc. CHI 2005*.
21. Nielsen, J. Heuristic evaluation. In Nielsen, J., and Mack, R.L. (Eds.), *Usability inspection methods*. John Wiley & Sons, New York, NY. (1994).
22. Pidgeon, N. Grounded Theory: theoretical background. In *Handbook of Qualitative Research and Methods for Psychology and the Social Sciences*. (1996).
23. Pidgeon, N. & Henwood, K. Grounded Theory: Practical Implementation. In *Handbook of Qualitative Research and Methods for Psychology and the Social Sciences*. (1996).
24. Sellen, A., Rogers, Y., Harper, R., & Rodden, T. Reflecting Human Values in the Digital Age. In *Comm. of ACM*. 52 (3). (2009).
25. Star, S. (1989). The structure of ill-structured solutions: boundary objects and heterogeneous distributed problem solving. In L. Gasser & M. Huhns (Eds.), *Distributed Artificial Intelligence*. Cited in Ackerman, M., & Halverson, C. (2000). Reexamining organizational memory. *Communications of the ACM*, 43(1), 58-64.
26. Strauss, A. & Corbin, J. *Basics of Qualitative Research*. (2nd Ed.) Sage Publications. (1998).
27. Thomas, G. & James, D. Reinventing Grounded Theory: Some questions about theory, grounded and discovery. *British Educational Research Journal*, 32 (6), (2006), 767-795.
28. Urquhart, C. An Encounter with Grounded Theory: tackling the practical and philosophical issues. In *Qualitative research in IS: issues and trends*. (2001).
29. Wixon, D. & Wilson, C. The Usability Engineering Framework for Product Design and Evaluation. In Helander, M., Landauer, T., and Prabhu, P. (Eds.) *Handbook of Human Computer Interaction* (2nd Ed.). Elsevier Science. (1997).
30. Wright, P. & McCarthy, J. Empathy and Experience in HCI. *Proc. CHI 2008*. 637-64