# Studying Design Cognition in the Real World Using the 'In Vivo' Methodology

Bo T. Christensen

Copenhagen Business School

Linden J. Ball

University of Central Lancashire

## INTRODUCTION

Much traditional research on design cognition has employed artificial settings and quasi-realistic tasks. For example, the pioneering studies of design by Eastman (1970) focused on highly constrained space-planning problems in which participants had to optimize the location of furniture items within a room of specified dimensions. Eastman's approach seemed to set the tone for design research for the next couple of decades, and although the tasks that researchers employed became more akin to genuine design problems they nevertheless only really 'imitated' key aspects of the problems that arise in professional design practice (see Cross, 2001, for a review of this early literature). What this restricted task focus meant, of course, was that much of what is important about the situated, contextualized and collaborative nature of design was inevitably omitted from the research agenda for many years, whereas it could well be that these social and cultural factors are paramount to understanding the authentic nature of design cognition. Over the past decade or so there has, thankfully, been a burgeoning of research studying design cognition 'in the wild', focusing on expert designers and collaborative design teams working on real design tasks in their natural environments. This move towards understanding design as it arises 'in vivo' is very much to be welcomed. Nonetheless, such research brings with it a great deal of complexity, which in turn necessitates the selection and use of very rigorous methods to achieve effective data collection and reliable data analysis. In this chapter we tackle head-on the methodological challenges that pertain to studying real-world design cognition.

Studies attempting to examine design cognition – especially at the level of the individual designer – have most often employed methodologies such as questionnaires (e.g., Römer, Pache, Weißhahn, Lindemann, & Hacker, 2001), qualitative interviews (e.g., Cross & Clayburn Cross, 1996) and diary self-reports (e.g., Ball, Evans, & Dennis, 1994; Pedgley, 2007). The latter methodologies, however, all involve a major component of retrospective or anecdotal reporting, which brings with it a whole host of concerns relating to the reliability of the resulting evidence. In particular, retrospective reporting draws on cognitive mechanisms that are known to be highly susceptible to biases and distortions that arise from forgetting and confabulation (Wilson, 2002). In other words, when it comes to asking participants to recall past events the constructive and adaptive nature of memory means that they cannot be expected to have accurate recollections of what actually transpired during a process such as design problem solving, Furthermore, for many processes, especially those associated with expert performance as arises in professional design practice, people may never have had conscious access to key elements of their cognitive processing in the first place (Perkins, 1981). For example, research on cognition in science (Dunbar, 1997) has shown that the conscious reconstruction of the steps that led to a discovery did not include significant elements and mechanisms that were recorded by an observer who was present during the original discovery process.

It seems clear, then, that the methodologies used to study creative phenomena such as discovery, innovation and design should take into account people's poor memory for the steps and mechanisms involved in the process, as well as their inability to reconstruct veridically the associated events. Furthermore, retrospective studies often provide a highly filtered view of a participant's cognitive processes, making such studies problematic for investigating the full range of mechanisms that underpin design cognition. To address all of the limitations that derive from the deployment of retrospective research methods we contend that the optimal approach to investigate design cognition is to make use of 'live' or 'on-line' research techniques.

In fact, design cognition has used one particular on-line methodology fairly extensively (although often intermittently) over the past 30 years, that is, the technique of protocol analysis (e.g., Craig, 2001, Cross, 2001; Cross, Christiaans, & Dorst, 1996), which involves participants being instructed to 'think-aloud' while solving a design problem. Ericsson and Simon (1980, 1999) developed the core theoretical framework and methodological guidelines associated with think-aloud protocol analysis, arguing that obtaining such protocols from participants did not significantly interfere with the nature of task processing. Ericsson and Simon proposed that such protocols arise from the reporting by participants of the current contents of short-term memory (i.e., the information that is being heeded or attended to at any moment in time). As such, verbal protocols are viewed as having the capacity to reveal accurately important aspects the processes that underpin problem solving (see Ball, Evans, Dennis, & Ormerod, 1997, and Ormerod & Ball, 2007, for further detailed discussion of pertinent theoretical and methodological issues relating to protocol analysis in the context of design cognition).

Eastman (1970), in his studies of space planning and architecture appears to have been the first to conduct a protocol analysis of design cognition, and since that time protocol analysis has been used to study various design phenomena, such as goal analysis, fixation and attachment to concepts as well as the role of sketching, opportunism, and modal shifts in design (see Cross, 2001, for an informative review). In 1994 the second Delft Workshop was entitled 'Research in Design Thinking II – Analysing Design Activity' (Cross et al., 1996) and focused exclusively on the use of protocol analysis in design studies. For this workshop a number of internationally esteemed design researchers were asked to analyze the same verbal protocols that had been obtained from an empirical investigation of designers. The outcome of this workshop certainly gave protocol analysis a major boost as a highly respectable methodology within the design research community.

Despite the evident strengths of verbal protocol analysis as a design research method - and notwithstanding its burgeoning use in the contemporary design research - we note that it is not without limitations. In particular, protocol-based studies focus on single participants verbalizing concurrently while performing given tasks. In this type of study, the participants are presented with special 'think-aloud' instructions that require them to verbalize everything that is currently going through their minds while tackling the set task. These instructions force the participants to verbalize, and if they are silent for short periods of time the experimenter will remind them to 'Please, think aloud' or to 'Keep talking'. Research has shown that forcing subjects to verbalize during problem solving can interfere with performance and change cognitive behaviours (Davies, 1995; Lloyd, Lawson & Scott, 1995). Likewise, Schooler and Melcher (1995) showed that enforced elicitation of think-aloud protocols may not only interfere with non-verbal modalities in creative cognition, thereby leading to inaccurate reporting of what is arising, but may also be detrimental to the very creative process itself. In a number of experiments, Schooler, Ohlsson, and Brooks (1993) showed that think-aloud protocols apparently interfered with participants' abilities to solve insight problems. The results could not be explained merely with respect to the conscious effort necessary to perform verbal 'on-line' self-reports of cognitive processes. Somehow enforcing the think-aloud requirement interfered with (i.e., 'overshadowed') the creative processes necessary for attaining solutions to the insight problems. In sum, it seems that enforced verbalization is problematic in the study of at least some types of cognition.

We also note that the 'typical' protocol analysis method that is deployed in the study of design cognition employs a laboratory set-up that relies on the use of artificially constructed design tasks (Cross, 2001) that require a short time span for solution attainment (typically less than 2 hours) using participants (sometimes non-experts) working in isolation (see Ball & Ormerod, 2001a, 2001b, for detailed critiques of this approach). All of the aforementioned factors obviously contrast with real-world design, where design tasks are usually highly complex and may span months or even years. Moreover, in real-world design the contextual setting is typically social and team-based, whereas most protocol analysis studies use individuals working on their own; even protocol analysis studies that do use team-based interactions often utilize teams of strangers, depriving the designers of their prevailing social network and normal interaction partners. In the real world an expert designer will also normally work in a personally-tuned environment (e.g., their own office) with access to personalized tools (Craig, 2001). This is unlike the laboratory set-up, where designers are asked to function in an unfamiliar and sterile environment without their familiar tools. Since experts rely on external aids such as drawings and notes (Norman, 1998), it is important to incorporate such aids in the study of design cognition, rather than focusing on verbalizations alone (Chi, 1997). Furthermore, in experimental settings the experimenter frequently acts out the role of 'the client', but interaction between designer and the quasi-client is typically restricted to scripted and prefabricated responses to anticipated design questions, thus prohibiting more natural conversations and a meaningful image of a real client (Craig, 2001).

The experimental settings employed in the typical protocol analysis study of design cognition have been found to have a major influence on the resulting protocol data (Cross, Christiaans & Dorst, 1996). Thus, protocol analysis studies of design seem to cry out for more ecologically valid research. Taking this criticism of protocol analysis into account, one way forward is to study the creative process 'online' in other ways than through the use of enforced thinking-aloud conducted in the laboratory. Several theorists have argued that understanding situated behavior is essential for framing research on cognition (Hutchins, 1995; Suchman, 1987; Lave & Wenger, 1991), and it is therefore somewhat paradoxical that given the highly contextualized nature of design activity, research on design expertise has typically ignored the role of situational and social factors, instead focusing on conducting laboratory style investigations where such factors are controlled for. This paradox led Ball and Ormerod (2000a, 2000b) to call for more widespread use of 'cognitive ethnography' in the study of design cognition. In essence, cognitive ethnography involves the adoption of a subset of the features of traditional ethnographic approach, and the deliberate violation of other features, in the pursuit of cognitive research goals. Most importantly it involves observational specificity (as opposed to the intensity of a prototypical ethnography), and it places a strong emphasis on verifiability in terms of validating observations across observers, data-sets and research techniques. Below we will extoll the virtues of one particularly important and potentially influential approach to utilizing cognitive ethnography methods to study creative design processes as they occur 'online' in the real-world.

## 'IN VIVO' RESEARCH ON DESIGN

Dunbar (e.g., 1995, 1997, 2000, 2001; Dunbar & Blanchette, 2001) developed a methodology for studying cognitive processes arising in science, called the 'in vivo/in vitro' method. The name is borrowed from the biologist's vocabulary in relation to research within the biological sciences. For example, a virus can be examined both in the Petri dish ('in vitro') and when it infects a host organism ('in vivo'). Similarly, Dunbar proposes, the same cognitive processes can be examined both in the laboratory, using controlled experiments, and as they occur 'live' in the real-world. This allows the cognitive researcher to investigate a phenomenon in a naturalistic fashion, and then go back into the psychological laboratory and conduct controlled experiments on what has been identified in such naturalistic settings (Dunbar, 2001). In this way the methodology attempts to maintain both the ecological validity highlighted as essential by a number of researchers (e.g., Neisser, 1976, Hutchins, 1995), as well as the experimental rigor that is possible in the psychological laboratory.

The 'in vivo' part of the method makes use of so-called 'messy' data (Chi, 1997), which refers to such things as verbalizations, observations, videotapes and gestures studied in naturalistic contexts. The in vivo/in vitro approach has been used with success in studying expertise in scientific domains such as physics, fMRi research, and astronomy (e.g., Trafton, Trickett & Mintz, 2005; Trickett & Trafton, 2002; 2007; Trickett, Trafton & Schunn, 2000), as well as other domains of expertise, such as the Mars Rover Mission (Chan, Paletz & Schunn, 2012). The in vivo methodology has also been transferred to studies of design cognition and design expertise, being used, for example, to examine analogical reasoning, mental simulation and requirement-handling in design teams (Ball & Christensen, 2009; Ball, Onarheim, & Christensen, 2010; Christensen & Schunn, 2007, 2009) and the co-evolution of problems and solutions in design creativity (Wiltschnig, Christensen, & Ball, in press).

The in vivo methodology involves eliciting qualitative data using ethnographic techniques, particularly audio-visual recording of design dialogue, and subsequently running the data through a rigorous coding scheme involving both quantitative and qualitative analyses to inform theories of design cognition. The methodology entails several data-processing stages, including: transcribing all verbal utterances; segmenting verbalizations into particular grain sizes; developing a coding scheme; applying the coding scheme; quantifying resulting patterns, sequences and occurrences of coded behaviours; and conducting reliability and validity checks. This kind of rigorous data-analysis approach has its origins in verbal protocol analysis (Ericsson & Simon, 1999), but extends protocol analysis to the handling of real-world discourse. As such, it utilizes the natural dialogue between designers as the main unit of analysis, with no special instructions being provided to 'think aloud'. Designers simply work as normal, and their activities and utterances are video/audio-recorded.

In order to ensure observational specificity and verifiability of results (cf. Ball & Ormerod, 2000b), regular time points need to be identified where design cognition occurs. Furthermore, the time points need to be located in a group setting so as to ensure that natural design dialogue will take place. Dunbar (1995, 1997) discovered that in the domain of molecular biology, a suitable time point related to the regularly scheduled laboratory meetings that are held by many scientists, especially in the natural sciences. Lab meetings involve discussion between senior scientists and their postdoctoral researchers and PhD students, and Dunbar found that these lab meetings contained a range of cognitive activities, such as hypothesis generation, the proposal of new experiments and criticisms of existing ones, and sometimes the development of entirely new concepts. He found that these meetings "... provided a far more veridical and complete record of the evolution of ideas than other sources of information" (Dunbar, 2001). In sum, lab meetings were especially well suited as an object of study where science could be investigated in a naturalistic context.

An analogous object of study in design turns out to be small team-based design meetings or product development meetings that typically arise in professional design contexts. One example is the medical plastics project analyzed in Christensen and Schunn (2007), where a sub-group in a large product development project was followed over an 8 month period, recording their weekly group meetings. The sub-group focused on producing novel features of the product, and consisted of 5 core members representing several disciplinary functions. The activity at the meetings was team-based and included a suitable number of people (typically 4 to 6) to allow for meaningful interaction. Because the designers were talking out loud during these meetings there was an external record of thinking and reasoning.

Pilot studies in these sub-group product development meetings showed that the design activity taking place consisted of a broad cross-section of what characterizes design thinking and reasoning in general. The primary function of these sub-group product development meetings was the creative development of design artifacts, that is, the actual activity of creating and problem solving in collaboration. Such activity included brainstorming, concept development, design problem solving, planning data collection and the next steps of design process, testing and evaluating mock-ups and prototypes, sketching, experimenting, and engaging in discussions and knowledge exchange about end-users, production methods and the like. In the meetings observed, 6% of the time concerned off-task verbalizations (such as office gossip, jokes, banter between the designers), 3% was spent summarizing the findings of past meetings (usually at the beginning of the meeting), 3% was spent planning future meetings (typically at the end of the meeting), 10% concerned planning future data collection or experiments, and 78% of the meetings concerned design thinking and reasoning in the 'here-and-now'. Thus, the majority of the time spent on these meetings appeared to focus on design thinking and reasoning. These percentages are, of course, context specific, and will likely be somewhat different in different organizational situations, different design projects, or different phases of the design process. But they illustrate that this particular methodology is promising as far as the analysis of design thinking and reasoning is concerned, in that it captures relatively little in the way of irrelevant data and is likely to generalize effectively to other studies of design cognition. Furthermore, these percentages illustrate that it may be beneficial to conduct tests of how much irrelevant data one is likely to capture given a particular object of study, so as to avoid drowning in irrelevant data. Reassuringly, however, the types of product development meetings analyzed by Christensen and Schunn (2007) appear to be very much the norm in design situations.

## DATA COLLECTION

'In vivo' research requires a great deal of background knowledge of the domain in question, since the data obtained derive from experts who are reasoning about their usual tasks. It is necessary, therefore, for the researcher to develop knowledge of the basic vocabulary and structure of the task, in order to understand what is going on (Ball & Ormerod, 2000a). As such, prior to commencing data collection, it is advisable for the researcher to become familiar with the design domain in question, through interviews and pilot studies as well as by reading background information relating to the organization, the design domain and the product type in question. Prior to each recorded sub-group meeting, it is also advisable for the researcher to conduct an interview with one of the designers so as to find out what the status of the project is, what is going to be the topics of the day's meeting, and what the group is currently working on – including any design difficulties they are experiencing.

When video- or audio-taping design team meetings at regular intervals, it is important to try to limit the inevitable loss of information about what has been going on since the last recording. A pre-meeting interview may help ensure that less information is lost. The meeting can then be videotaped, and the conversation between the designers audio-taped. When recording in vivo there appears to be a trade-off between the amount of data that can be collected and the invasiveness of the data collection procedure. The procedure can, of course, potentially influence the process if designers become too self-conscious or uneasy about the situation when being recorded. A fairly non-invasive method is audio-taping, although this omits a great deal of potentially important information about design objects that are present, people's motor activities, gestures and emotional expressions and the eye-gaze of the designers.

A much more invasive method for collecting some of these potentially important data can involve multiple cameras set to record the total-room view as well as desktops, the gestures of individual designers and details of any note-taking or sketching behavior. Such an approach is likely to influence the behavior of the designers unless care is taking to hide all recording equipment as much as possible and allow for long trial periods to allow the participants to adapt to the artificial feel of the situation. Much design activity can be captured using a single camera that is located above and a short distance away from the table where the designers are situated during the meeting, but zoomed in so that all objects can be discerned, and all sketching and note-taking activities can be captured, albeit not in detail. This allows for all people to be in the frame and enables the subsequent examination of who was currently talking if this cannot be discerned from the audiotape. Bodily gestures and general gaze can also be discernible to some extent, depending on the bodily posture of the individual designers (e.g., gaze cannot be perfectly discerned when looking away from the camera). Using such a camera view, facial expressions are probably not so easily discernible.

In terms of audio-recording an omnibus microphone (i.e., a microphone recording sound from all directions simultaneously) linked to the videotape can be placed at the center of the room or table to allow for the recording of all verbalizations. No special instructions (e.g., to ‘think-aloud’) should be given to participants at the meeting; they should simply be informed that they should proceed with the meeting as they normally would. An observer responsible for collecting data can make written notes of any information not readily available in the video frame and can also gather any handouts or documents. Following each meeting all design objects (e.g., sketches, mock-ups and prototypes) present during the meeting can be videotaped in close-up, sometimes with one of the designers explaining in voice-over the function of the object. This voice-over, together with the videotape, provides valuable information regarding what design object (e.g., sketch, sketch part or prototype) was currently being referenced in the verbalizations arising during the design meeting.

## DATA ANALYSIS

Following data collection all verbalizations are transcribed. Once transcribed the data can then be analyzed as a series of statements, using standard verbal protocol analysis techniques (e.g., Ericsson & Simon, 1999). These statements can potentially reveal a lot of detail about the cognitive mechanisms that are operating during creative reasoning processes, as Dunbar has shown (e.g., Blanchette & Dunbar, 2000, 2001; Dunbar, 1997, 1999, 2001). The transcription process is time-consuming, and typically takes 7 to 10 hours of labor per hour of video/audio. The transcribed data can then be segmented (divided into units) according to a suitable grain size, with typical units being propositions, sentences or episodes (i.e., statements linked contemporaneously to a common theme or goal). For much design reasoning research a useful grain size involves dividing the data in terms of ‘complete thought’ segments (e.g., Hughes & Parkes, 2003). This entails separating verbal statements into segments containing verb phrases which are indicative of mental operations. Each segment will typically be either a single sentence or a fragment of a sentence, yielding hundreds of segments per hour of transcript. Each segment can be given a time stamp and additional non-verbal codes can be added to the segments if desirable. For example, eye gaze details, gestural information and referenced objects can all be coded from the video data.

*Data reduction*

Initially, the data-set might be reduced by applying preliminary codes that focus in on the relevant parts of transcripts. For example, applying a code for ‘off-task’ as opposed to ‘on-task’ verbal behavior can remove irrelevant passages where the designers engage in office-related banter, personal gossip or making jokes – or any other verbalizations that are not related to the task at hand. Another example is that transcripts can be divided into episodes. An episode is a chunk of segments that share a common theme (e.g., they all concern planning the next meeting, or they all deal with evaluating a particular prototype). By dividing transcripts into episodes, certain types of episodes can be excluded from further coding insofar as they are irrelevant to the hypotheses being tested. Great care should obviously be taken in selecting episodes for exclusion from subsequent data analysis, since this could potentially raise doubts as to whether the chosen sub-set of data is a valid representation of the remainder of the transcript.

*Coding schemes*

In order to test hypotheses and theories of design thinking and reasoning, coding schemes have to be developed. The development of a coding scheme is difficult to convey in general terms, since this depends heavily on the researcher’s theoretical orientation, the hypotheses or questions being asked, the task and the domain (Chi, 1997). The reader is referred to Ericsson and Simon (1999) for further details regarding the development of effective coding schemes. However, a few illustrations are provided below.

*Coding schemes for cognitive processes: Mental simulation*

Creative cognitive processes, such as analogical transfer and mental simulation, can be captured using qualitative screening of sequences of segments. Here the researcher’s interest lies in the cognitive processes involved in design. An illustrative example is the code for mental simulations that was adapted from Trickett and Trafton’s (2002, 2007) coding scheme of scientists running mental models during data analysis. A mental model run is a mentally constructed model of a situation, phenomenon or object that can be grounded in memory or in a mental modification of the design objects currently present. This allows the designers to think and reason about new possible states of the design object and its perceptual qualities, features and functionality, without actually having physically to change the object.

The key feature in a mental simulation is that it involves a simulation ‘run’ that alters the representation in order to produce a change of state. This means that the simulation is not merely a question that is asked (e.g., changing features or functions of the design object); it also provides a kind of answer (e.g., Will this design work? How should this design be produced?). Mental simulations thus represent a specific sequence that starts with the creation of an *initial representation*, and which then moves to *running* the representation, whereby it is modified by spatial transformations (e.g., where elements or functions are extended, added or deleted), followed lastly by a *changed representation*. These three elements (initial representation, run and changed representation) are not mutually exclusive and can occur within the same segment, although frequently they will cover several segments. The code for mental simulation has been applied in analysing in vivo data deriving from design contexts in a number of publications (e.g., Ball & Christensen, 2009; Christensen & Schunn, 2009; Ball, Onarheim, & Christensen, 2010; Wiltschnig, Christensen, & Ball, in press). An example of a mental simulation is presented in Table 1.

Table 1. An example of a mental simulation

| | |
|---|---|
| Initial representation | Could you add something so that you couldn't close this thing because there would be something in the way when you try to fold this way… |
| Run | But if this thing goes this way, then it is in a position to allow the ear to enter... But then I just don't know how it should be folded… 'cause if it is folded this way then it will come out here…then it should be folded unevenly somehow…You should fold it oblique. |
| Changed representation | It wouldn't make any difference one way or the other. It would fold the same way, and come out on this side the same way. |

The mental simulation code is a qualitative code, which makes it quite time-consuming to apply to transcripts since there is no quick way of identifying the occurrence of such simulations. The coders must code each segment in turn, noting elements of mental simulations as they go along. Furthermore, the code requires that the coder understands much of the context for each segment. Past research has yielded very high inter-rater reliability for this code.

*Coding schemes for mental states: Epistemic uncertainty*

An example of a mental state of interest in design cognition is ‘epistemic uncertainty’. Epistemic uncertainty concerns a reference to one’s own subjective experience of uncertainty, that is, a mental state of feeling uncertain or lacking adequate knowledge or understanding. One way to code for epistemic uncertainty is to use a purely syntactical approach. This approach has been employed by Trickett, Trafton, Saner, and Schunn (2005) in an analysis of expert meteorologists and fMRI researchers performing spatial transformations. Trickett et al. used hedge words to locate segments displaying uncertainty, with these hedge words including terms such as ‘probably’, ‘sort of’, ‘guess’, ‘maybe’, ‘possibly’, ‘don’t know’, ‘[don’t] think’, ‘[not] certain’, ‘believe’ and so on. Segments containing these hedge words were located and coded as ‘uncertainty present’ if a scrutiny of the individual segment confirmed that the hedge word concerned uncertainty. The coding scheme for epistemic uncertainty has also been applied successfully in design contexts (Christensen & Schunn, 2009; Ball & Christensen, 2009; Ball, Onarheim, & Christensen, 2010; Wiltschnig, Christensen, & Ball, in press). Syntactical codes are relatively easy to apply (see Table 2 for examples), but they are only suitable for a limited number of categories.

Table 2. Examples of epistemic uncertainty coded syntactically using hedge words

| Utterance | Code |
|---|---|
| ’Cause I’m **not sure** whether you would fold it around the back. | Uncertain |
| I **think** so too, but before we get too cocky, let’s make a model… | Uncertain |
| Well, I **guess** it’s a combination of moist and heat isn’t it? I **suppose** it has to be. | Uncertain |

| | |
|---|---|
| It has to push from the start | Not uncertain |
| Yes, but the problem is that you can't hit it later …'cause it's too small | Not uncertain |
| It...then we have...then we lose the possibility of folding it back. | Not uncertain |

*Coding schemes for episodes: Co-evolution of problem and solution*

Arguably one of the most important theoretical concepts in research on design cognition – and certainly one of the most cited notions – concerns the idea that problems and solutions 'co-evolve' during the design process (Maher, 1994; Dorst & Cross, 2001). Under this view, design concepts are often viewed as developing iteratively, with the design problem and associated solutions co-evolving in a mutually adaptive manner. Wiltschnig, Christensen, and Ball (in press) developed a code for identifying co-evolution episodes. This first of all entailed coding for 'requirement mentions', which involved examining each segment to determine whether it contained an explicit reference to a design requirement and whether this requirement mention concerned: (a) adding a novel requirement; (b) interpreting or making revisions to an existing requirement; (c) bracketing a requirement (as in "We're not going to be dealing with that here"); or (d) deleting a requirement. In order for co-evolution to be deemed to be taking place, requirements had to be linked to solution attempts, such that each segment was further coded for whether the designers suggested a solution or an idea in order to fulfill the design requirements. Finally, co-evolution episodes were coded by linking requirement mentions to solution attempts and by looking qualitatively at the content of each. Insofar as a requirement mention was categorized as being linked to a solution attempt within a 5-minute transcript window, this was thence coded as being a 'co-evolution episode'.

Table 3. An example of a co-evolution episode

| | |
|---|---|
| A | ... Could we make it so that it opens itself ... so you do not actually need to go down and fiddle with the flap here |
| B | Well there are some people who want to ... |
| C | There are some people who want to bend the flap over backwards |
| A | Yes ... backwards… perhaps we should have a look at the hygiene around this flap. |
| B | Yes, but here we can write uhh |
| A | Yes increased hygiene on uhh |
| B | improvements ... improved cleaning of the bottom flap or what? |
| D | There uh there an injection-moulded solution is again a good idea because it will be easier to wipe clean. |

A theory-driven choice in relation to the construction of an a priori coding scheme may end up being too general for straightforward application to particular verbal data. What this means is that once a coding scheme has been developed, a decision then has to be made as to which verbalizations constitute evidence that they can be translated into a particular code. In other words, the coding should be *operationalized* in relation to the context and type of data at hand. For example, if one wants to study differences in analogical distance between different analogies that are present in the transcript, it is one thing to have a general, theoretically interesting distinction between ‘local’ analogies and ‘distant’ analogies, and quite another to know how to code for this distinction in the current data-set. In his research on molecular biologists, for example, Dunbar (1995, 1997) operationalized this local/distant distinction by creating three categories: 'within organism', 'between organism' and 'non-biological or distant' analogies. For the operationalization of analogical distance in a design context see Christensen and Schunn (2007), Ahmed and Christensen (2009) and Ball and Christensen (2009).

*Reliability and validity*

The nature of in vivo data also requires that the researcher pays particular attention to reliability analyses. Reliability is important in any methodology that is used for studying design cognition, but may be particularly important in relation to in vivo data because of the somewhat high degree of contextual variance that can arise, as opposed to the relative contextual stability in experimental settings. Inter-rater reliability checks of individual codes using independent coders can be conducted using Cohen's Kappa statistic rather than the mere percentage agreement that some researchers have reported. Percentage agreement will make agreement seem much higher than warranted, especially when locating phenomena that are relatively rare ('needle-in-a-haystack') in a large data set. Since this is often the case in in vivo data, even an exceedingly high percentage agreement can be problematic.

A satisfactory level of inter-rater reliability using Cohen's Kappa depends on whether the codes are equiprobable and on the number of codes applied, but as a general rule of thumb, Fleiss (1981) suggested that below 0.40 is characterized as poor agreement, 0.40 to 0.75 as fair to good agreement, and above 0.75 as excellent agreement (see also Landis & Koch, 1977, for another way of labeling the magnitude of agreement). Other types of reliabilities are also important; for example, when possible it is a good idea to recode the same hypotheses using a different coding scheme and grain size (assumed to tap into the same hypotheses), to see if the in vivo results hold up (e.g., Chi,

1997) . Furthermore, the reliability of the results may be illustrated through splitting the dataset, and showing that the results are stable across the individual parts of the data-set.

Due to the extensive data analysis and coding involved when undertaking in vivo research, the method will typically involve the analysis of only a relatively few hours of recordings. Furthermore, for the same reasons, usually a rather small number of different contexts are studied. This limited data variance and data quantity can potentially threaten the generalizability of the results due to an increased risk of sampling error and low-N problems. Therefore, as mentioned, Dunbar recommends supplementing in vivo research with in vitro controlled experiments that can better deal with these sampling and low-N issues. These limitations aside, in vivo research remains particularly well suited for tackling the lack of ecological validity in much design cognition research.

*Using the same data-set to answer different research questions*

Because in-vivo data are collected without an experimental setup, the same data-set can be used to answer different research questions using alternative techniques such as discourse analysis or conversation analysis. This was evident, for example, in papers presented as part of the Seventh Design Thinking Research Symposium (DTRS7; McDonnell & Lloyd, 2009), where a single data-set of design dialogue was analysed by several independent research teams using distinct analytic approaches to investigate a range of unique research questions.

## CONCLUSION

The in vivo methodology holds much promise as a means to improve on some of the key limitations of more traditional methodologies that have been deployed in studying design cognition. In vivo research attempts to study design thinking and reasoning 'live' and 'online' as it actually takes place in the real world. In relation to engineering design contexts we have argued in this chapter that sub-group product development meetings may be suitable objects of study, with pilot studies of multidisciplinary design teams revealing that the verbal content arising in sub-group meetings reflects a broad cross-section of design activities, with a majority of the time being take up with verbalizations that are indicative of highly engaged processes of design thinking and reasoning. By recording the verbalizations that arise in such meetings and subsequently transcribing, segmenting and coding the data, we have shown in this chapter that it is possible to test specific hypotheses about the authentic nature of design cognition as it arises in the real-world.

In contrast to more traditional design research methodologies, the in vivo approach has some major advantages. First and foremost, the methodology captures design thinking and reasoning 'live' as it actually occurs, in contrast to many other methodologies that focus on retrospective data, which is known to be highly biased and inaccurate in nature. Furthermore, although in vivo research shares many of the data analysis features of think-aloud protocol analysis it avoids the problems that arise from having an enforced verbalization requirement, which is typically a standard aspect of verbal protocol studies. Rather, in vivo research relies on the *natural* dialogue that takes place spontaneously between designers. In addition, whilst a typical protocol-based study takes place in an experimental laboratory setting, in vivo research focuses on real-world design with expert designers working on their normal tasks, in their usual contexts, using their personalized tools, collaborating with their regular networks and teams and developing their design ideas over extensive periods of time. This all serves to ensure that in vivo design research has much better ecological validity than standard experimental and protocol-based design research.

However, in vivo design is not without some problems. It can be labor intensive in terms of data analysis and coding issues, which may put the approach at risk of sampling errors and low-N issues if an overly restricted number of cases are subjected to analysis. To reduce this potential threat to the generalizability of the results, it is recommended that in vivo research is supplemented with standard experimental lab-based studies that can add rigor by significantly increasing the number of analyzed cases.

We contend that all of the issues with the in vivo methodology are surmountable, especially if it is viewed as one of a set of techniques within a 'methodological triangulation' approach that focused on clarifying the validity, generalisability and reliability of emerging findings. The burgeoning use of the in vivo method in investigating issues such as the roles of mental simulation, analogising, epistemic uncertainty and problem—solution co-evolution in creative design augurs well for its on-going deployment in future design research. As experienced users of the in vivo approach ourselves we commend the methodology to other design researchers who are committed to analysing the *authentic* nature of design cognition as it arises in real-world collaborative contexts.

**REFERENCES**

Ball, L. J., & Christensen, B. T. (2009). Analogical reasoning and mental simulation in design: Two strategies linked to uncertainty resolution. *Design Studies, 30*, 169-186.

Ball, L. J., Evans, J. St. B. T., & Dennis, I. (1994) Cognitive processes in engineering design: A longitudinal study. *Ergonomics, 37*, 1753-1786.

Ball, L. J., Evans, J. St. B. T., Dennis, I., & Ormerod, T.C. (1997) Problem-solving strategies and expertise in engineering design. *Thinking & Reasoning, 3,* 247-270.

Ball, L. J., Onarheim, B., & Christensen, B. T. (2010). Design requirements, epistemic uncertainty and solution development strategies in software design. *Design Studies, 31,* 567-589.

Ball, L.J., & Ormerod, T.C. (2000a). Applying ethnography in the analysis and support of expertise in engineering design. *Design Studies 21*, 403-421.

Ball, L.J., & Ormerod, T.C. (2000b). Putting ethnography to work: The case for a cognitive ethnography of design. *International Journal of Human-Computer Studies 53*, 147-168.

Blanchette, I. & Dunbar, K. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory & Cognition 28*, 108-124.

Blanchette, I., & Dunbar, K. (2001). Analogy use in naturalistic settings: The influence of audience, emotion, and goals. *Memory & Cognition 29*, 730-735.

Chan, J., Paletz, S. B. F., & Schunn, C. D. (2012). Analogy as a strategy for supporting complex problem solving under uncertainty. *Memory & Cognition, 40,* 1352-1365.

Chi, M. T. H. (1997). Quantifying qualitative analysis as verbal data: A practical guide. *The Journal of the Learning Sciences 6*, 271-315.

Christensen, B. T. & Schunn, C. D. (2007). The relationship of analogical distance to analogical function and pre-inventive structure: The case of engineering design. *Memory & Cognition, 35*, 29-38.

Christensen, B. T., & Schunn, C. D. (2009). The role and impact of mental simulation in design. *Applied Cognitive Psychology, 23*, 327-344.

Craig, D. L. (2001). Stalking *Homo Faber*: A comparison of research strategies for studying design behavior. In C. M. Eastman, W. M. McCracken, & W. C. Newstetter (eds.) *Design knowing and learning: Cognition in design education* (pp. 13-36). Amsterdam: Elsevier.

Cross, N. (2001). Design cognition: Results from protocol and other empirical studies of design activity. In C. M. Eastman, W. M. McCracken, & W. C. Newstetter (eds.) *Design knowing and learning: Cognition in design education* (pp. 79-104). Amsterdam: Elsevier.

Cross, N., Christiaans, H., & Dorst, K. (1996). Introduction: The Delft Protocols Workshop. In N. Cross, H. Christiaans, & K. Dorst (eds.) *Analysing design activity* (pp. 1-16). Chichester, UK: John Wiley & Sons.

Cross, N., & Clayburn Cross, A. (1996). Winning by design: The methods of Gordon Murray, racing car designer. *Design Studies, 17,* 91-107.

Davies, S. P. (1995). Effects of concurrent verbalization on design problem solving. *Design Studies 16*, 102-116.

Dorst, K., & Cross, N. (2001). Creativity in the design process: Co-evolution of problem—solution. *Design Studies, 22*, 425-437.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (eds.) *The nature of insight* (pp. 365-395). Cambridge, MA: MIT Press.

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. In T.B. Ward, S.M. Smith, & J. Vaid (eds.) *Creative thought: An investigation of conceptual structures and processes* (pp. 461-493). Washington, DC: American Psychological Association.

Dunbar, K. (1999). How scientists build models: In vivo science as a window on the scientific mind. In L. Magnani, N. Nersessian, & P. Thagard (eds.) *Model-based reasoning in scientific discovery* (pp. 89-98). New York, NY: Plenum Press.

Dunbar, K. (2000). How scientists think in the real world: Implications for science education. *Journal of Applied Developmental Psychology 21*, 1, 49-58.

Dunbar, K. (2001). The analogical paradox: Why analogy is so easy in naturalistic settings yet so difficult in the psychological laboratory. In D. Gentner, K.J. Holyoak, & B.N. Kokinov (eds.)

*The analogical mind: Perspectives from cognitive science* (pp. 313-334). Cambridge, MA: MIT Press.

Dunbar, K. (2001). What scientific thinking reveals about the nature of cognition. In K. Crowley, C.D. Schunn, and T. Okada (eds.) *Designing for science: Implications from everyday, classroom, and professional settings.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Dunbar, K., & Blanchette, I. (2001). The in vivo/in vitro approach to cognition: The case of analogy. *Trends in Cognitive Sciences 5*, 334-339.

Eastman, C. M. (1970). On the analysis of intuitive design processes. In G. T. Moore (eds.) *Emerging methods in environmental design and planning*. Cambridge, MA: MIT Press.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review 87*, 215-251.

Ericsson, K. A., & Simon, H. A. (1999). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York, NY: John Wiley.

Hughes, J., & Parkes, S. (2003). Trends in the use of verbal protocol analysis in software engineering research. *Behaviour & Information Technology 22*, 127-140.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159-174.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Lloyd, P., Lawson, B., & Scott, P. (1995). Can concurrent verbalization reveal design cognition? *Design Studies 16*, 237-259.

Maher, M. L. (1994). Creative design using a genetic algorithm. *Computing in Civil Engineering*, *2*, 2014-2021.

McDonnell, J., & Lloyd, P. (2009). (Eds.). *About: Designing. Analysing design meetings.* London: Taylor & Francis.

Neisser, U. (1976). *Cognition and reality*. San Francisco, CA: W. H. Freeman and Co.

Norman, D. A. (1988). *The psychology of everyday things*. New York, NY: Basic Books, Inc.

Ormerod, T. C., & Ball, L. J. (2007). Qualitative methods in cognitive psychology. In C. Willig & W. Stainton-Rogers (Eds.), *Handbook of qualitative research in psychology* (pp. 553-574). London: Sage Publications Ltd.

Pedgley, O. (2007). Capturing and analysing own design activity. *Design Studies, 28,* 463-483.

Perkins, D. N. (1981). *The mind's best work*. Cambridge, MA: Harvard University Press.

Römer, A., Pache, M., Weißhahn, G., Lindemann, U., & Hacker, W. (2001). Effort-saving product representations in design—results of a questionnaire survey. *Design Studies, 22*, 471-491.

Schooler, J. W., & Melcher, J. (1995). The ineffability of insight. In S. M. Smith, T. B. Ward, & R. A. Finke (eds.) *The creative cognition approach* (pp. 97-133). Cambridge, MA: MIT Press.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122*, 166-183.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication.* New York, NY: Cambridge University Press.

Trafton, J. G., Trickett, S. B., & Mintz, F. E. (2005). Overlaying images: Spatial transformations of complex visualizations. *Foundations of Science*.

Trickett, S. B., & Trafton, J. G. (2007). 'What if': The use of conceptual simulations in scientific reasoning. *Cognitive Science, 31,* 843–875.

Trickett, S. B., Trafton, J. G., Saner, L. D., & Schunn, C. D. (2005). I don't know what's going on there: The use of spatial transformations to deal with and resolve uncertainty in complex visualizations. In M. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 65-86). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Trickett, S. B., & Trafton, J. G. (2002). The instantiation and use of conceptual simulations in evaluating hypotheses: Movies-in-the-mind in scientific reasoning. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 878-883). Mahwah, NJ: Erlbaum.

Trickett, S .B., Trafton, J. G., & Schunn, C. D. (2000). Blobs, Dipsy-Doodles and other funky things: Framework anomalies in exploratory data analysis. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 965-970). Mahwah, NJ: Erlbaum.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: The Belknap Press.

Wiltschnig, S., Christensen, B. T., & Ball. L. J. (in press). Collaborative problem—solution co-evolution in creative design. *Design Studies*.