A methodology for studying design cognition in the real world

The in vivo research methodology holds promise to improve some of the limitations of typical design cognition methodologies. Whereas typical design cognition methodology use protocol analysis (utilizing special 'think-aloud' instructions and/or artificial settings) or retrospective analyses, in vivo research attempts to study design thinking and reasoning 'live' or 'online' as it takes place in the real world. No special instructions are used since the method relies on natural dialogue taking place between designers. By recording verbalizations at product development meetings (or other suitable objects of study), transcribing, and coding the data, it is possible to test hypotheses about design cognition in the real-world. This promises to improve the ecological validity over typical design cognition studies. Problems with the methodology include labor-intensiveness leading to small samples (possible sampling errors). To deal with this problem, it is recommended to supplement in vivo research with traditional larger sample laboratory studies.

Bo T. Christensen
Department of Psychology
University of Aarhus
botc@psy.au.dk

INTRODUCTION: DESIGN RESEARCH METHOD LIMITATIONS

Design activity includes cognitive processes such as problem solving and creativity [9], making the design domain an obvious choice (along with science and art) when cognitive scientists want to explore higher cognitive functions. Although it has been proposed that design problem solving may differ in some respects from other kinds of problem solving, the distinctions are not always sharp enough to warrant a domain-independent theory of design problem solving [43]. This led Zimring & Craig [43] to argue for a design research á la carte, treating design problem solving as problem solving in general, and to focus on the reasoning processes involved (such as analogy, mental simulation, argumentation, decision making, synthesis) as these processes help construct novel and useful artifacts. Such research could potentially contribute to theoretical development in cognitive science, and facilitate the development of problem solving and creativity research that can cross narrow design disciplinary boundaries.

Studies attempting to examine design cognition usually employ methodologies such as protocol analysis, questionnaires, and interviews. Retrospective or anecdotal evidence (such as historical analyses, interviews, dairy studies or questionnaires about design processes) from designers has been used to try to pinpoint the cognitive mechanisms behind design cognition.

These retrospective methods are however very unreliable when dealing with cognitive mechanisms where the subject cannot be expected to have accurate memory of – or perhaps even conscious access to - what exactly is going on in the process [32]. For example, research on cognition in science has shown that conscious reconstruction of the steps that led to a discovery did not include significant elements and mechanisms that were recorded by a present observer [12]. Subjects' poor memory of the steps and mechanisms involved in creative processes, as well as their inability to accurately reconstruct the events, should be taken into account in the methodology used to study such phenomena. Further, retrospective studies often provide a highly filtered view of the subjects cognitive processes, making them problematic in studies of the processes and mechanisms in design cognition. Therefore, it has been recommended that design cognition be studied using 'live' or 'on-line' research methods.

In fact, design cognition has used a particular 'on-line' methodology (protocol analysis) in the past 30 years [7], where subjects are instructed to 'think-aloud' while solving design problems. The use of protocol analysis seems to have increased in recent years. Ericsson & Simon [19,20] developed think-aloud protocols, and argued that they did not significantly interfere with, and could accurately report the content of short term memory, and thereby reveal the processes going on in regular problem solving.

Eastman [18] studying architecture was the first to conduct a protocol analysis in design, and since that time protocol analysis has been used to study for example, goal analysis, co-evolution of problem and solution, fixation and attachment to concepts, the role of sketching, opportunism, and modal shifts [see 8 for a review]. In 1994 the second Delft Workshop was entitled 'Research in Design Thinking II – Analysing Design Activity' [9] and focused exactly on protocol analysis. Here a number of design researchers were asked to analyze the same verbal protocols derived from experimental studies of designers. The outcome of this workshop appears to have given protocol analysis a boost in the design literature.

But even the 'online' methodology of protocol analysis is problematic as a methodology to study design cognition. A major part of protocol analysis studies focus on single subjects verbalizing concurrently while performing a given task. In this type of study, the subjects are given special 'think-aloud' instructions to verbalize all that is currently going through their head while performing the task. These instructions force the subjects to verbalize, and if they grow quite for short periods of time, the experimenter will remind them to 'please, think aloud' or 'keep talking'. Recent research has shown that forcing subjects to verbalize during problem solving can interfere with performance or change cognitive behavior [10,29]. Schooler et al. showed that not only may forced 'think-aloud' protocols be inaccurate in reporting what is going on in creative cognition by interfering with non-verbal modalities [33], but they are also detrimental to the very creative process they seek to study [34]. In a number of experiments, Schooler et al. [34] showed that think-aloud protocols apparently interfered with subjects' abilities to solve insight problems. The results could not be explained merely with respect to the conscious effort necessary to perform verbal 'online' self-reports of cognitive processes. Somehow forced think-aloud protocols interfered with ('overshadowed') the creative processes going on. Thus it seems that forced verbalizations are problematic in the study of at least some types of cognition.

Further, the typical protocol analysis study employs an experimental laboratory setup using relatively simple and artificially constructed design tasks [8] with a very short time span (typically less than 2 hours) using subjects (sometimes non-experts) working on their own [2]. This obviously contrasts with real-world design where the typical design task is highly complex and may span months or years. In real world design the contextual setting is typically social and team-based, but most protocol analysis studies use individuals working on their own, and even protocol analysis studies using team-based interactions often utilize teams of strangers, depriving the designers of their persistent social network and normal interaction partners. In the real world, the individual expert designers work in a personally tuned environment (e.g., their own office) with personalized tools etc [7], unlike the laboratory where they are asked to function without such tools. Since experts rely on external aids such as drawings and notes [31], it becomes increasingly important to incorporate such aids and tools in the study of design cognition, rather than focusing on verbalizations alone [5]. Further, in experimental settings the experimenter is frequently used as 'the client', but interaction between designer and 'client' is restricted to scripted and prefabricated responses to anticipated design questions thus prohibiting more natural conversations and a meaningful image of the client [7]. These experimental settings employed in the typical protocol analysis study have been found to have a heavy influence on the protocol data [9]. In contrast, several theorists have argued that understanding situated behavior is essential for framing research on cognition [27,36,28], and it is somewhat paradoxical that given the highly contextualized nature of design activity, research on design expertise have typically ignored the role of situational and social factors to conduct laboratory style investigations where such factors are controlled for. This led [2,1] to call for an applied or cognitive ethnography in the study of design cognition. Thus, protocol analysis studies of design seem to cry out for more ecologically valid research about how the design process takes place in the real-world.

Taking this criticism of protocol analysis into account, it is necessary to study the creative process 'online' in other ways than through forced 'think-aloud' protocols conducted in the laboratory. One such 'online' methodology would be to study the creative process, as it occurs 'live' in the real-world. Dunbar [e.g., 11,12,14,16,17] has recently created a methodology for studying cognition in science, called the in vivo-in vitro method. The name is borrowed from the biologist's vocabulary on biology research. 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 the naturalistic settings [16]. This way, the methodology attempts to maintain both the ecological validity highlighted as essential by a number of researchers [e.g., 30,27,6], as well as the experimental rigor that is possible in the psychological laboratory. In vivo research makes use of so-called messy data [5], which refers to such things as verbalizations, observations, videotapes and gestures studied in naturalistic settings. The in vivo – in vitro approach has been used with success in studying expertise in scientific domains such as physics, fMRi research, and astronomy [37,38,39,40,42], as well as other domains of expertise, such as meteorology and submarine operations [41]. Until now the methodology has not been applied in the study of design. But recently the present author has used this methodology to study engineering design cognition, and below I will focus on the in vivo part of this methodology, and show how it can be used to study design cognition while avoiding some of the limitations and pitfalls of the usual design cognition methodologies.

IN VIVO RESEARCH ON DESIGN

The present version of in vivo research was constructed to study design cognition – notably thinking and reasoning - as it takes place in naturalistic design situations amongst expert practicing engineering designers. I identified a major international company working in medical plastics who had shown consistent design skill and creativity over a number of years. The product development department had won multiple design awards. The company agreed to take part in this study, and I was given access to the company and all aspects relating to a particular design project that was about to start up (spanning more than 2 years), including interviews with members of the project, access to the product database, email correspondence, access to meetings at all levels, including brainstorming meetings, observations of end-user product evaluation sessions, decision making meetings at both the micro and macro level, and more. I followed the design project for the first 8 months (primarily the concept design phase) of the design project. Initially a number of interviews were conducted to familiarize myself with the company and the way the project I would be following was organized. The goal was to identify points in time where creative design thinking occurs and capture this on audio or video tapes that could then be analyzed for the processes involved in the thinking and reasoning in design cognition. The time points in question would preferably be recurrent on a regular basis (e.g., occurring at regular times every week) and contain a cross-section of design activities, so as to allow for the study of multiple different design activities, allow for analysis of development over time (i.e., development across different time points as the design process progressed), and allow for the practical issue that I could schedule attendance to these time points in advance, rather than having to be present at the company at all times, as would frequently be the case in ethnographic studies. Further, the time point should be set in a group setting to ensure that natural dialogue would take place. Dunbar [11,12] had discovered that in the domain of molecular biology, a suitable time point was the regularly scheduled laboratory meetings held by many scientists, especially in the natural sciences. Lab meetings consist of a senior scientist along with his or her Post Doc.s and PhD students, and Dunbar found that lab meetings contain a range of cognitive activities, such as hypothesis generation, proposal of new experiments and criticism 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" [16]. This made the lab meetings well suited as an object of study where science could be studied in a naturalistic context.

An analogous object of study in engineering design turned out to be product development meetings. The design project I would be studying incorporated 19 people who were loose or permanent members of the project. This large group was also organized into smaller units focusing on different aspects of the overall design. For example, one such sub-group focused on producing completely novel features of the product, and consisted of 5 core members (representing multidisciplinary functions, e.g., engineering, architecture, production). This subgroup (like all the subgroups involved in the project) held weekly product development meetings. Because the designers were talking out loud there was an external record of thinking and reasoning. Thus by recording product development meetings it is possible to gain access to 'online' thinking and reasoning without influencing the way the designers think. Using this method it is possible to directly monitor thinking and reasoning rather than uncovering reasoning through post-hoc interviews, questionnaires or think aloud protocols [12]. Pilot studies in these subgroup product development meetings showed that the design activity taking place in these groups consisted of a broad cross-section of what characterizes design thinking and reasoning in general. The primary function of these subgroup product development meetings were creative development of design artifacts – that is actual creating and problem solving in collaboration – and the activity included brainstorming, concept development, design problem solving, planning of data collection and the next steps of design process, testing and evaluating mock-ups and prototypes, sketching activity, experiments, discussions and knowledge exchange about end-users, production methods and more.

A concern when conducting in vivo research is that because such research takes place in a naturalistic environment, it is likely that large amounts of irrelevant data will be captured. A risk facing the in vivo researcher is that of drowning in irrelevant data. Unlike artificial experimental settings, where the experimenter actively sets up a very particular task and context to study a particular phenomenon, in vivo research has to try to locate a suitable object of study in the real world. This approach is likely to capture irrelevant data that has to be weeded out during an often quite extensive data collection and data analysis work load. To reduce this concern of capturing too much irrelevant data, it is important to pick the object of study carefully, so that the captured irrelevant data can be kept to a minimum, and drowning in data can be avoided. To examine whether the subgroup product development meetings primarily concerned design cognition, I coded for content in these subgroup product development meetings to find out how much of the time was allocated to design thinking and reasoning. The average results indicated that, in the meetings I observed, 6% of the time concerned off-task verbalizations (such as office gossip, jokes, banter between the designers), 3% were spent summarizing the findings of past meetings (usually at the beginning of the meeting), 3% were spent planning future meetings (typically at the end of the meeting), 10% concerned planning future data collection or experiments, and a full 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. Note that these exact percentages would probably be somewhat different in a different organizational context, different design project, or different phase of the design process than I studied. These percentages are merely to illustrate that this particular object of study is promising in the study of design thinking and reasoning, in that it captures relatively little irrelevant data, and looking for similar objects of study in other design projects holds promise. Further, it is meant to 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.

Besides these subgroup product development meetings, other types of meetings were also held, carrying different functions. For example, leaders of each subgroup would meet with the head of the design project on a bimonthly basis to discuss strategy and status of the project. But pilot studies of these strategy meetings revealed that the content of these meetings to a much lesser extent focused on thinking and reasoning about creating design artifacts. The subgroup product development meetings were thus selected as a highly suitable object of study. It's content including a broad cross section of design activities in general, design thinking and reasoning occupied the vast majority of the time at these meetings, the regularity of the meetings, and the activity was team based and included a suitable number of people (typically 4-6) to allow for meaningful interaction. These types of product development meetings appear to be somewhat typical in engineering design teams, and there is no reason to assume that this highly suitable object of study is special for the organization or design project I was studying.

DATA COLLECTION

In vivo research requires a great deal of background knowledge of the domain in question, since the data involves experts thinking and reasoning about their usual tasks. Therefore it is necessary for the researcher to develop knowledge of the basic vocabulary and structure of the task, in order to understand what is going on. Therefore I conducted interviews with members of the subgroup as well as the project leader to familiarize myself both with the type of design product in question, the organization of the project and subgroup, the nature and steps of the design process about to begin and so on. Further, I read information about existing products in the same domain, sat in on strategy and decision making meetings, conducted pilot studies and in other ways familiarized myself with the domain and typical design process, and the vocabulary and habits of the designers that I could expect to encounter. Following this initial data gathering and familiarization, I started to collect data on the object of study (the subgroup product development meetings).

Prior to each meeting, I conducted a semi-structured interview with one of the designers to find out what the status of the project was, what was going to be the topics of today's meeting, and what they were currently working on, along with any design problems they were having. I then attended the meeting as an observer only. The meeting was videotaped, and the conversation between the designers was audio taped. When recording in vivo there appears to be a tradeoff between amount of data that can be collected, and the invasiveness of the data collection procedure (that can potentially influence the process if the designers become too self-conscious or stressed of being recorded). A non-invasive method is audio-taping only, which obviously lacks a lot of potentially important information about design objects present, motor activities and gestures, gaze of the designers etc, while capturing only verbalizations. A highly invasive method collecting some of this potentially important data could involve multiple cameras set to record total-room view, desk-tops, gestures of individual designers, and details of any note-taking or sketching behavior. Such an approach will probably 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. I chose to collect an amount of data that would be relatively non-invasive, while still collecting most of the important variables. A single camera was set up high above and a short distance from the table where the designers sat during the meeting, but zoomed in so that all objects on the table could be discerned, and all sketching and note-taking activities could be captured, albeit not in detail. All people present were in the frame to allow for an examination of who was currently talking if this could not be discerned on the audiotape. Bodily gestures and general gaze could be discerned most but not all the time, depending on bodily posture of the individual designers (e.g., gaze could not be perfectly discerned when looking away from the camera). Facial expressions could not be discerned. An omnibus microphone linked to the videotape was placed center table to allow for recording of all verbalizations.

No special instructions (e.g., instructions to ‘think-aloud’) were given to participants at the meeting – they were simply informed that they should proceed with the meeting as they normally would. As an observer I took notes of information not readily available in the video frame, and collected any handouts. Following each meeting all mock-up and prototypes that had been present during the meeting was videotaped in close-up, sometimes with one of the designers explaining in voice-over the function of the object. Sketches were also recorded or copied when possible. This, together with the videotape, allowed for noting what design object (e.g., sketch, prototype or part of sketch) was currently being referenced in the verbalizations. The interviews and additional information gathered provided supplemental sources of information. The primary object of study was the videotaped meetings.

DATA ANALYSIS

Following data collection all verbalizations are transcribed. Once transcribed, the data can then be analyzed as a series of statements following standard verbal protocol analysis fashion [20]. These statements can potentially reveal a lot about the cognitive mechanisms operating during the creative and reasoning processes, as Dunbar [3,4,12,13,15] has shown. The transcription process is time consuming, and typically takes 7-10 hours labor per hour of video/audio.

The transcribed data can then be segmented (divided into units) according to a suitable grain size (i.e., size of each segment, such as proposition, sentence, episode). For much design thinking and reasoning research, such a grain size could be dividing the data into ‘complete thought’ segments [e.g., 26]. 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 fraction of a sentence, yielding hundreds of segments per hour of recording. Each segment can then be given a time stamp, and additional non-verbal codes can be added to segments if necessary (e.g., gaze, gestures, referenced object, and so on can be coded from the video data). These segments are the primary unit of analysis.

In order to test hypotheses and theories of design thinking and reasoning, a coding scheme has to be developed. It is very difficult to convey the steps involved in choosing specific codes, since it depends entirely on the researcher’s theoretical orientation, the hypotheses or questions being asked, the task and domain [5]. Developing and operationalizing a coding scheme is a task too complex to be described here in detail, but in essence this coding scheme development procedure follows standard verbal protocol analysis, and the reader is referred to Ericsson & Simon [20] for more details. Rather, I will provide an extended example from ongoing research on engineering design illustrating different types of codes in the next section. Choosing a coding scheme should be done a priori so as to reduce the chance that post hoc theory will influence the data [20], but the first theory-laden choice of coding scheme may be too general for application on particular verbal data. Therefore, once having chosen a coding scheme, it needs to be decided what verbalizations in the data constitute evidence that they can be translated into a particular code. In other words, the codings 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 in the verbalizations, 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 a particular data set. In molecular biology, Dunbar [11,12] operationalized this distinction by creating three categories: ‘within organism’, ‘between organism’ and ‘non-biological or distant’ analogies.

A few general comments of special relevance to in vivo data should be noted here. First, in vivo data is typically much less specific than data collected under artificial constraints in the experimental laboratory. This means that somewhat large amounts of irrelevant data will be present – even when care is taking in selecting relevant objects of study. This irrelevant data can be weeded out by applying preliminary codes that focuses 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 talk office and personal gossip, make jokes, banter, and other verbalizations 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, in so far as they are irrelevant to the hypotheses being tested. But obviously care should be taken in selecting episodes for exclusion from data analysis, since this could potentially raise doubts about whether the chosen subset of data is a valid representation of the remainder of the transcripts. The nature of in vivo data further requires that the researcher pay particular attention to reliability analyses. Reliability is important in any methodology studying design cognition of course, but may be particularly important in in vivo data because of the somewhat high degree of contextual variance (as opposed to the relative contextual stability in experimental settings). Inter-rater reliability checks of individual codes using independent coders should be conducted using Cohen’s Kappa measures rather than the mere percent agreement that some researchers have reported. Percent 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 percent agreement can be problematic. A satisfactory level of inter-rater reliability using Cohen’s Kappa should be above

.70. 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 holds up [e.g., 5]. As can be gathered by the above description, the data analysis and coding part of in vivo research is extremely labor intensive.

EXAMPLE: MENTAL SIMULATION AND UNCERTAINTY IN REAL-WORLD DESIGN

The move from hypotheses to coding scheme is difficult to describe in general terms and the issue is too complex to deal with in this conference proceeding. Further, this part of in vivo research is not much different from standard verbal protocol studies, and so the reader is referred to Ericsson & Simon [20] for more details. So instead of describing the process in general terms, I will offer a concrete example of a coding scheme using a few different types of codes from my own data of engineering designers. These codes are from work in progress and the codes have been selected for illustrative purposes, meaning that the background of the hypotheses and the results are not explained in detail. The main focus here is on the move from hypothesis to coding scheme, and on providing examples of codes that can be utilized.

The hypothesis to be tested, deals with the relation between information uncertainty and mental models. A mental model is a representation of some domain or situation that supports understanding, reasoning, and prediction [23]. Mental models rely on qualitative relationships, such as signs and ordinal relationships, and relative positions [e.g., 22]. Relevant to engineering design, mental models have been used to explain human reasoning about physical systems, including devices and mechanisms [35,25,24]. An important feature of mental models is that they frequently permit mental simulation. A mental simulation refers to the sense of being able to dynamically ‘run’ a simulation internally to observe functioning and outcome of a system or device. ‘Runnability’ implies a sense of being able to simulate system behavior and predict outcomes even for situations where the subject has no previous experience. This has been termed ‘mental simulation’, ‘mental model runs’ [23] and ‘conceptual simulation’ [38] – and here these terms are used synonymously.

Mental model runs have some disadvantages as a thinking strategy, notably inaccuracy and imprecision [23]. However, the potential advantage of using mental model runs in design include being able to reason about how physical systems will operate under changed circumstances/with altered features, without having to resort to actually physically constructing such a system or device. This implies quick and cheap ways of testing possible alternatives. This is particularly useful in creative domains, such as science, art and design, where uncertainty is an inescapable part of the problem space since the task involves constructing novelty. Constructing novelty implies moving into the beforehand unknown possibilities and impossibilities of the subject matter [6]. There are multiple ways of attempting to deal with the inherent uncertainty in design, including experimentation and other data collection, analogical thinking, and the actual construction of objects – but mental model runs may be yet another way. Mental model runs may help in the reasoning and thinking about such possibilities and impossibilities, thus reducing some of the uncertainty associated with design. Some support for this had been found in the domain of science, where use of mental models has been linked to information uncertainty and ambiguity. Trickett [40,38] found that the majority of mental simulations in scientific data analysis was used to evaluate hypotheses (i.e., an areas of scientific thinking fraught with uncertainty), and argued that mental model runs were used as a strategy to help resolve uncertainty. Mental simulations were used as frequently as or more frequently than any other strategy, and thus played a significant role in scientists’ consideration and evaluation of hypotheses.

The present analysis was an attempt to extend this hypothesis into the domain of engineering design to see if it would hold up under different circumstances. The hypothesis being tested was thus whether information uncertainty leads to mental model runs as an attempt to reduce this uncertainty. The constructs to be measured are thus ‘information uncertainty’ and ‘mental simulations’.

The engineering design transcripts used as data were 9 hours of video taken from the data collection described above as ‘subgroup product development meetings’ in the product development department of a major company in medical plastics. All 9 hours of data were from the same subgroup. These 9 hours of video were transcribed and segmented according to complete thought. The segmentation produced a total of 7414 segments covering 7 different transcripts. Added to the transcripts were information about design objects present at the meeting to ease the coding of which design objects were currently being referenced in the protocols.

A coding scheme was developed to first limit the data set to product development in the here-and-now (i.e., reduce the transcripts to include only relevant segments), and second, to code for information uncertainty and mental model runs (we will primarily focus on the second step). The first step of the coding included coding for off-task verbalizations, segments dealing with planning future meetings or data collection, and segments dealing with referencing past meetings. The percentages of the transcript of each of these codes were reported above. This left 78% or 5806 segments of on-task here-and-now design thinking and reasoning.

The second step involved coding for information uncertainty and mental simulation.

Mental simulation

The code for mental simulations were adapted from Trickett’s [40,38] 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 to physically change the object. But mental simulations do not just concern the technical aspects of the design object, but can include a host of other types of simulations of changed circumstance. One frequently occurring type concerned simulating contextual shifts, such as end-user behavior and preferences under changed circumstances (e.g., using a novel design object). The key feature in a mental simulation is that it involves a simulation ‘run’ that alters the representation, to produce a change of state [38]. This means that the simulation is not merely a question asked (e.g., changing features or functions of the design object); it also provides a kind of answer (e.g., will it work, how should it be produced). Mental simulations thus represent a specific sequence starting with creating an initial representation, running the representation (it is modified by spatial transformation where elements or functions are for example 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 in the same utterance/segment, although frequently they will cover several segments. Each segment was coded as ‘mental simulation’ (1) or ‘no mental simulation’ (0).

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 some how…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.

Example of a mental simulation

The mental simulation code is a qualitative code making it quite time-consuming since there is no quick and dirty way of identifying mental simulations in a transcript. The coders must code each segment in turn, noting elements of mental simulations as they go along. Further, the code requires that the coder understands much of the context for each segment, meaning that it is necessary to know about design and about what is being developed in this particular design transcript.

But past research has yielded high inter-rater reliability for this code.

Information uncertainty

To illustrate different kinds of codes, two different measures of information uncertainty will be used, one relying on syntax, the other on a combination of verbal and visual information taken from the video.

Information uncertainty using syntax. One way to code for uncertainty is to use a purely syntactical approach. This approach was adapted from [41] who used hedge words to locate segments displaying uncertainty. These hedge words included for example words like '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' (1) if a scrutiny of the individual segment confirmed that the hedge word concerned uncertainty.

All other on-task here-and-now segments were coded as 'no uncertainty' (0). Syntactical codes are quite easy to apply, but they can only be applied to a limited number of categories.

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 its too small	Not uncertain
It...then we have...then we loose the possibility of folding it back.	Not uncertain

Examples of information uncertainty using syntax

Information uncertainty using verbalizations and video in combination. A different way of approaching the measure of information uncertainty is to look specifically at the objects of design thinking, or 'pre-inventive structures' [21]. These objects can take many different forms, including prototypes, sketches, mock-ups, or simply be ideas that are unsupported by external representations (neither in 3d physical form or on paper). It could be argued that these different kinds of design objects have different degrees of information uncertainty, in that they represent different levels of specification of the concept in question. In that line of thinking, an idea left unsupported by sketches or prototypes is more 'uncertain' than the prototype where technical features and functions are much more specified. Ideas, sketches and prototypes are all 'ambiguous' in a general sense in that they can be reinterpreted and changed somewhat rapidly, and in the sense that they represent an object-in-the-making, rather than a finished form. But the ambiguity and uncertainty may be somewhat less for prototypes than for ideas, with sketches somewhere in between. Sketches primarily support visual representation but is less specified in other modalities (haptic, gustatory, olfactory, and auditory). Therefore in design, we would expect that experts working with external support systems of sketching and prototypes would be facing less artifact uncertainty, than when no such external support exists ('idea only'). Further, sketching would provide more uncertainty than prototypes. Thus, another way of measuring information uncertainty is to code for the kind of design object being referenced. In the present transcript three different kinds of design objects occurred frequently: Prototypes, sketches and 'ideas' (i.e., objects of design thinking that were unsupported by external representation). This distinction is referred to as 'type of preinventive structure' below. Included in the transcripts were information about the design objects present at each meeting (sketches, prototypes etc.). For each segment it was first coded whether the focus of attention of the person speaking was one of these design objects present in the room. This was coded using the video recording of the design session (not the verbal data). Focus of attention was operationalized as either actual handling or holding a particular object; pointing to a particular object; or gazing toward a particular object (if this was possible to discern from the video). In effect the 'focus of attention' code acted as a helping variable in coding type of preinventive structure. Then coders coded whether each segment of the verbal data referred to an 'idea' (1), sketch (2), prototype (3) or other (4-removed from analysis), aided by the 'focus of attention' variable. Note it is of course perfectly possible to look or handle one type of object and think about another. In all cases the verbalized objects had precedent, meaning that if there was a difference between referenced object between focus of attention and verbalization, the object from the verbalization was chosen. Coding the preinventive structure variable was quite time consuming given that both video analysis and then verbal protocol analysis were required, but ‘focus of attention’ from the video data greatly aided the coding of the verbal protocols since most of the segments had a synchronicity between focus of attention and verbal reference.

Reliability

Following coding various forms of reliability were conducted. Inter-rater reliability was done on 17% of the data (two full transcripts), with all disagreements resolved by discussion. All inter-rater reliability tests reached satisfactory Kappa values. The syntactical uncertainty measure and the mental simulation codes both had exceedingly high Kappa values (>.90). Further, two split-half reliability analyses were conducted to test for ordering effects. Each transcript was split in half, and all analyses were re-done using the first halves and second halves separately. The transcripts were then rank ordered in terms of data collection date, and the first half of the transcripts were separated from the last half of the transcripts, and analyses were re-done on each of these halves. All split-half reliability tests yielded comparable results.

Results

The results revealed that mental simulations were extremely common in engineering design.

Chi-squares analyses revealed that segments containing syntactical information uncertainty had significantly more mental simulations than segments without uncertainty, supporting the hypothesis that information uncertainty and mental model runs are linked. Another chi-square showed significant differences between ideas, sketches and prototypes. Subsequent 2x2 chi-squares revealed that idea and sketches did not differ, but both had significantly more mental model runs than prototypes. These results converge to lend support to the hypotheses that a link exists between information uncertainty and mental model runs in real-world engineering design. The link was strong enough to show up using two different codes for uncertainty under naturalistic circumstances in real-world design, thus demonstrating a strong and psychologically meaningful effect. However, since these results are correlational in nature we cannot draw any firm conclusions as to causality. We thus need more research before we can conclude that mental model runs are used *as a strategy* to reduce information uncertainty in design. The present results suffer from possible sampling biases in that only a small number of sessions and subjects were involved. More research both in vivo and in vitro should be conducted to replicate these findings.

An important advantage in using in vivo research is the fact that the data is not collected in order to test one particular hypothesis, but rather can be used to test a range of hypotheses. The nature of the data collection allows for an infinite number of re-codings of the transcribed data. These re-codings can concern finer grained analyses of the same or similar hypotheses using different codes. But the same set can also be used again in testing *other* hypotheses about thinking and reasoning in concept design, for example concerning analogical thinking, aesthethics, design planning and so on. In the domain of science this can be illustrated in the works of Susan Trickett and colleagues. They collected data on scientific data analysis in the domains of physics, astronomy and cognitive psychology, and used *the same* data sets to analyze hypotheses about conceptual simulation when evaluating hypotheses [40], anomalies in data analysis [42], and change of representation in visual data analysis [39].

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

CONCLUSION

The in vivo methodology holds promise to improve on some of the limitations of typical design cognition methodologies. In vivo research attempts to study design thinking and reasoning ‘live’ or ‘online’ as it takes place in the real world. In engineering design, it is argued that subgroup product development meetings may be suitable objects of study, in that pilot studies in multidisciplinary design teams reveal that the content includes a broad cross section of design activities in general, and because design thinking and reasoning occupies the majority of the time at these meetings. By recording verbalizations at such meetings (or other suitable objects of study) in the real world, transcribing, segmenting and coding the data, it is possible to test hypotheses about design cognition in the real-world. In contrast to more traditional design methodologies, this approach has some advantages. In vivo methodology captures design thinking and reasoning ‘live’ as it occurs, as contrasted with some design methodologies focusing on problematic retrospective data. Further, although in vivo research sharesmuch of the data analysis features of protocol analysis it avoids the problematic forced verbalizations typically used in verbal protocol studies. Rather, in vivo research relies on natural dialogue taking place between designers. While the typical protocol study takes place in an experimental lab setting, in vivo research focuses on real world design with expert designers working on their normal tasks, in their usual context and using personalized tools, working with their regular network and teams, over extensive periods of time. This ensures that in vivo design research will prove to have a much better ecological validity than standard experimental and protocol design research. However, in vivo design is not without problems. It can be somewhat problematic due to the labor intensive data analysis and coding issues, which may put in vivo research at risk of sampling errors and low N problems, if too few 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 studies that may add experimental rigor and significantly increase the number of analyzed cases.

REFERENCES

1. Ball, L.J. and Ormerod, T.C. Applying ethnography in the analysis and support of expertise in engineering design. *Design Studies 21*, 4 (2000), 403-421.
2. Ball, L.J. and Ormerod, T.C. Putting ethnography to work: The case for a cognitive ethnography of design. *International Journal of Human-Computer Studies 53*, 1 (2000), 147-168.
3. Blanchette, I. and Dunbar, K. How analogies are generated: The roles of structural and superficial similarity. *Memory and Cognition 28:* (2000), 108-124.

4. Blanchette, I. and Dunbar, K. Analogy use in naturalistic settings: The influence of audience, emotion, and goals. *Memory and Cognition 29*, 5 (2001), 730-735.

5. Chi, M.T.H. Quantifying qualitative analysis as verbal data: A practical guide. *The Journal of the Learning Sciences 6*, 3 (1997), 271-315.

6. [Removed to preserve anonymity in blind review]

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

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

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

10. Davies, S.P. Effects of concurrent verbalization on design problem solving. *Design Studies 16*, 1 (1995), 102-116.

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

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

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

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

15. Dunbar, K. The analogical paradox: Why analogy is so easy in naturalistic settings yet so difficult in the psychological laboratory. In D. Gentner, K.J. Holyoak, and B.N. Kokinov (eds.) *The analogical mind: Perspectives from cognitive science.* The MIT Press, Cambridge, MA (2001), 313-334.

16. Dunbar, K. 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.* Lawrence Erlbaum Associates,Inc.,Publishers, Mahwah, N.J., US (2001)

17. Dunbar, K. and Blanchette, I. The invivo/invitro approach to cognition: the case of analogy. *Trends in Cognitive Sciences 5*, (2001), 334-339.

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

19. Ericsson, K.A. and Simon, H.A. Verbal reports as data. *Psychological Review 87,* 3 (1980), 215-251.

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

21. Finke, R.A., Ward, T.B., and Smith, S.M. *Creative cognition: Theory, research, and applications*. MIT Press, Cambridge, MA (1992).

22. Forbus, K.D. and Gentner, D. Qualitative mental models: Simulations or memories? In *Proceedings of the Eleventh International Workshop on Qualitative Reasoning,* (1997)

23. Gentner, D. Psychology of mental models. In N.J. Smelser and P.B. Bates (eds.) *International encyclopedia of the social and behavioral sciences*. Elsevier, Amsterdam (2002), 9683-9687.

24. Hegarty, M. Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition 18*, 5 (1992), 1084-1102.

25. Hegarty, M. and Just, M.A. Constructing mental models of machines from text and diagrams. *Journal of Memory & Language 32*, (1993), 717-742.

26. Hughes, J. and Parkes, S. Trends in the use of verbal protocol analysis in software engineering research. *Behaviour & Information Technology 22*, 2 (2003), 127-140.

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

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

29. Lloyd, P., Lawson, B., and Scott, P. Can concurrent verbalization reveal design cognition? *Design Studies 16*, 2 (1995), 237-259.

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

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

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

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

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

35. Schwartz, D.L. and Black, J.B. Analog imagery in mental model reasoning: Depictive models. *Cognitive Psychology 30*, (1996), 154-219.

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

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

38. Trickett, S.B. Movies-in-the-mind: The instantiation and use of conceptual simulations in scientific reasoning. Unpublished doctoral dissertation from George Mason University (2004)

39. Trickett, S.B., Fu, W.-T., Schunn, C.D., and Trafton, J.G. From dipsy-doodles to streaming motions: Changes in representation in the analysis of visual scientific data. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society.* Erlbaum, Mahwah, NJ (2000), 959-964.

40. Trickett, S.B. and Trafton, J.G. 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*. Erlbaum, Mahwah, NJ (2002), 878-883.

41. Trickett, S.B., Trafton, J.G., Saner, L., and Schunn, C.D. 'I don't know what's going on there': The use of spatial transformations to deal with and resolve uncertainty in complex visualizations. (In press)

42. Trickett, S.B., Trafton, J.G., and Schunn, C.D. 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. Erlbaum, Mahwah, NJ (2000), 965-970.

43. Zimring, C. and Craig, D.L. Defining design between domains: An argument for design research á la Carte. In C.M. Eastman, W.M. McCracken, and W.C. Newstetter (eds.) *Design knowing and learning: Cognition in design education*. Elsevier, Amsterdam (2001), 125-146.