

# Getting access to what goes on in people's heads?

## - Reflections on the think-aloud technique

Janni Nielsen  
Torkil Clemmensen  
Carsten Yssing

Department of Informatics  
Copenhagen Business School  
Howitzvej 60, 2000 Frederiksberg, DK  
+45 3815 2400

[jn.inf@cbs.dk](mailto:jn.inf@cbs.dk), [tc.inf@cbs.dk](mailto:tc.inf@cbs.dk), [cy.inf@cbs.dk](mailto:cy.inf@cbs.dk)

### ABSTRACT

One of the basic usability testing techniques the HCI community draws on, and which stands out as unique, is thinking aloud. We introduce the many names, uses and modifications of the classical think aloud technique, and ask the rhetorical question: What do researchers think they get when they ask people to think aloud? We answer it by discussing the classical work of Ericsson and Simon(1984), in particular their distinction between vocalisation, verbalisation and retrospective reports and the relation to short term memory. Reintroducing the psychological perspective and the focus on higher order cognitive processes, we argue that access to subjective experience is possible in terms of introspection and describe a technique that invites the user to become a participant in the analysis of his or her own cognitive processes. We suggest that use of think aloud has as a prerequisite explicit descriptions of design, test procedure and framework for analysis. We point out, however, that if the aim is to get access to human thinking, HCI research may benefit from experimental research.

### Keywords

Think aloud, verbal report, usability.

### INTRODUCTION

Usability testing and evaluating systems involves the testing of users' interaction with the computer, and the literature on the techniques applied is vast. Although the actual form and use may vary, some of the basic techniques employed by the HCI community are: Think aloud, observations, video recordings, automatic logging of cursor movements or keystrokes, guided interaction, interviewing, and questionnaires. There is an understanding within the field that testing, using only one technique, is insufficient. Therefore, many of the

techniques mentioned above are combined. However, in usability testing the thinking-aloud technique stands out as special.

Thinking aloud was originally described by Karl Duncker(1945) in his work within experimental psychology where he studied productive thinking. In HCI, researchers often position their discussion of techniques in relation to think aloud (Nielsen J. 1992, Waes 1998; Nielsen & Christiansen 2000; Buur & Bagger 1999; Karsenty 2001), however there are only few references to the original work of Duncker within usability research.<sup>1 2</sup> The think aloud technique is frequently mentioned as having been applied – though it is not necessarily described in detail nor discussed (Koenemann-Belliveau et al. 1994; Rowley 1994; Bringham, John & Lewis 1991; Yeo 1998). In HCI practice, thinking aloud seems to be one of the most popular techniques. It is often referred to as the usability method and used both in laboratory settings, workshops and field testing (Kensing 1998; Nielsen 1992; Rowley 1994). In a survey of methods and techniques used by HCI practitioners in Denmark (main body of respondents) and researchers (about 25% of the respondents), thinking aloud appeared to be the single most frequently applied technique in testing (Clemmensen 2002). This should not come as a surprise – internationally the technique is taught as part of the HCI curriculum at many universities and described in many

<sup>1</sup> The use of verbal protocols is not limited to usability testing. There is a large amount of literature on the use of the technique to study writing, reading strategies, text comprehension and decision-making, and the technique has a long tradition in clinical psychology.

<sup>2</sup> We are aware that Clayton Lewis wrote one of the first papers on the topic (Lewis, C. "Using the thinking-aloud method in cognitive interface design," IBM Research Report RC 9265, Yorktown Heights, NY, 1982), but we have been unable to obtain a copy of the paper.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page.

To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

NordiCHI 10/02 Århus, Denmark

© 2002 ACM ISBN 1-1-58113-616-1/02/0010...\$5.00

textbooks. Dix, Finlay, Abowd and Beale(1997) credit think-aloud for its simplicity and advocate a more relaxed view of the process, pointing out that the usefulness of think aloud is seen as "...largely dependent on the effectiveness of the recording method and subsequent analysis" (p. 427). Molich (1994) sees think-aloud as a technique that is ready to use with proper handling, and Hackos and Redish (1998) also regard it as a very straightforward technique. They suggest asking the user to think aloud while conducting contextual inquiry site visits. "This is the same technique that we use during usability testing and the purpose is the same – to get users' inferences, intuitions, and mental models ...reasons...decisions...while doing the task" (p. 137). "By recording the verbal protocol, you will be able to ...detect cognitive activities that may not be visible at all" (p. 259). Especially Jakob Nielsen (1992, 1994) has been tireless in promoting the technique and its benefits. However, Preece (1994) points to the cognitive load and added strain on users as well as the interruptive role of the observer during the think aloud test. Preece, Rogers and Sharp (2002) also comment that the user probably would find it difficult to speak when the task became demanding. The whole situation would probably feel awkward. Silence is the likely outcome of the situation and also the biggest problem. A solution that is recommended suggests having two people work together – talk to each other. Yet, the technique is tempting because it promises instant results and seems to be cost effective, because only few tests subjects are needed. The technique can even be used by non-usability specialists (Nielsen J. 1994; Wright & Monk 1990). But most important – it is assumed to give access to the cognitive processes during users' engagement with computers.

Many efforts are used and high value placed on user testing. However, it is interesting that there seems to be a lack of research literature reflecting on users' application of the technique (Branch 2000)<sup>3</sup>. How do users experience it? What do users think of it? Teaching graduate students in Informatics to think aloud and asking them to reflect on their experience with using the technique have raised a number of issues. Students complain that they think faster than they can speak, that their thought processes are much more complex than they can verbalise, and that thinking aloud interferes with their interaction with the interfaces and the task. Besides, thinking aloud does not come naturally to most people. Although the students took part in a course on HCI tools and techniques and were familiar with user testing, and

<sup>3</sup> Jennifer Branch asked adolescents who were participating in her study on information seeking behaviour which technique they preferred: think aloud or think after. Although there were differences in preference, Branch discards this as there were "no differences in the extent to which participant spoke freely and openly in the study" (p. 384).

despite explicit assurance that it was the system and not them that were being tested, they reported, once they became test subjects, that they felt they were being observed, evaluated and judged and that it influenced their performance. If the interaction with the interfaces repeatedly broke down, they would often explain and excuse it by saying that they were stupid, did not really understand the task or more creatively suggest that they were unable to read the screen properly because they "needed new glasses".

The paper is organised as follows. We start by introducing some of the many names that thinking aloud is called by. We combine it with a presentation of some of the many uses and we point out that behind these often lay modifications, redesign and extensions of the classical think aloud technique – without the rationale for the changes nor reflection on the consequences of the changes being explicitly stated. Hence, we ask the rhetorical question: what do researchers think they get when they ask people to think aloud. We answer the question by introducing the readers to the classical reference to think aloud, the publication by Ericsson and Simon on verbal protocols (1984). We discuss their distinction between three levels of verbalisations: vocalisation, verbalisation and retrospective reports, including their relation to short term memory. We reflect critically on their assumption about a one-to-one relation embedded in the reduction of verbal protocols to "pure data". And we point out that Ericsson and Simon substituted Nisbett and Wilson's (1977) psychological perspective on verbal reports for an information processing perspective. We reintroduce the notion of introspection and point out possible ways to improve access to what goes on in peoples head. We argue that the thinking aloud technique requires explicit descriptions of design, test procedure and framework for analysis. However, if the aim is to get access to human thinking, we suggest that experimental research is included in the research agenda, arguing that the Scandinavian research tradition may benefit from this.

#### **A dear child with many names and many uses**

In the literature, thinking aloud is described under many names: verbal reports, concurrent verbal protocols, retrospective verbal protocols, after think aloud and verbal protocols. Even the name think-aloud sometimes seems to embed an uncertainty: the concept is placed in brackets: "think aloud" without any explanations as to why (Koenemann-Belliveau et al. 1994). This is even the case in the title of a paper (Katalin 2000). In papers reporting on studies using thinking aloud, it is often stated that the technique applied is different, which seems to mean different from the classical think aloud. However, the authors do not discuss and detail what they did, nor reflect on the technique. As shown by Boren and Ramey (2000) in a study on think aloud, the classical thinking aloud technique by Ericsson and Simon is seldom applied, but often redesigned, modified or extended with other techniques. Buur and Bagger(1999) redesigned it into a

user dialogue - to enhance “a dialogue between designers and users on use, context and technology” (p. 63) and to develop “...workshops where think-aloud was changed into “co-discovery learning””(p. 65). In their study on designers making their own evaluations, Wright and Monk (1990) explain how they applied the think-aloud technique, arguing that their version of the technique involved the user directly in the design. Users were told to think of themselves as “co-evaluators” of the system. They were asked questions such as “what will the system do if...?” and “Why did you do that?”. Hence, the role of the facilitator in the thinking aloud sessions was to ask hypothetical questions that would facilitate speculation about the system and questions about the users’ actions. The facilitator was also instructed to question the user when he/she asked questions to acquire insight into users understanding. Van Waes (1998) combined thinking-aloud with on-line recording of mouse movements and browser actions. This was supplemented with a questionnaire prior to the test and interviews with participants about their experience after the test sessions. Nielsen and Christiansen (2000) combined video recordings of thinking aloud sessions with a subsequent interviews where a priori identified video sequences were the object of a dialogue interview between researcher and user. By letting the actual events and the actual sequence of events organise the dialogue, they seem to get around some of the problems with retrospective interviews, which rely entirely on the users’ memory.

In an interesting study by Koenemann, Carroll, Rosson and Singley (1994), the aim is to follow up on the distinction between critical incidents and critical threads in usability testing. Interaction with the computer is seen as a process and not as isolated steps from breakdown to breakdown. The case involved students engaged in learning processes where the thinking aloud technique was applied. Videotapes and observation notes enabled a minute-by-minute description of learners’ activity captured in episode analysis. There is no description of how think aloud was introduced to the learners and no description of the test leaders’ role and action during the test. However, the transcripts included in the paper present glimpses of how the users sometimes speak in prescriptive comments: “I am about to make this”(p. 246), other times “talk” with the interaction: “I guess these are classes ... so ... I don’t know what a class is but I ...” (p. 249) and at other times reflect by analogy: “It’s kind of like hardware design ... when you build a piece of hardware you go to the store and buy this chip, and this chip, and this resistor...”(p. 247). With its minute-by-minute protocol, the approach to verbal reports seems to come close to the classical technique by Ericsson and Simon. However, Koenemann et al. were interested in all three kinds of verbal reporting, including the retrospective reports, whereas Ericsson and Simon were only interested in level one, vocalization, and level two, verbalization, but rejected the third level.

### Guidelines and models

Just like the literature on studies using think aloud is vast, so is the literature on guidelines. Guidelines on how to conduct field lab test using thinking aloud have been developed (Rowley 1994) as well as guidelines for teachers on how to conduct thinking aloud tests (Katalin 2000). Also, the number of subjects needed in thinking aloud tests has been discussed, and in a study of non-usability specialists using the think aloud technique, Jakob Nielsen (1994) suggests that five test subjects are sufficient to find the majority of problems. Discussing two different usability tests of online documentation - testing a prototype and testing a product after it was released - Mark Simpson (1990) introduces guidelines for testing methods to test online documentation iteratively. He suggests a general framework that starts by asking essential questions as a prerequisite to the planning. On this basis, he describes procedures and methods and suggests that verbal protocols tend to work best when a product is functional enough to be used for most of the tasks, for which it was intended.

Taking point of departure in designing usable multimedia online documentation, Mehlenbacher(1993) points out that the literature on guidelines for design and testing is immense. However, he argues that it lacks; “..how one goes about designing and implementing such tests, that is, how one selects a test group, designs usability materials, or decides which method provides what types of data?” (p. 211). Mehlenbacher tries to fill the gap by describing and discussing the strength and weaknesses of different techniques. Also here, thinking-aloud seems to be one of the foremost techniques, although he calls it talk-aloud. He does not talk of verbal protocols, but of talk-aloud protocols. Though the wording differs explicitly from the traditional use, the author fails to tell the readers what we should be aware of. However, going through his list of references, it seems that he is also taking his point of departure in the classical technique of verbal reports by Ericsson and Simon. Interestingly, he points out that the technique has been criticised for failing to elicit tacit knowledge – but he does not question that it provides valuable information in testing.

### What do you think you get when you ask somebody to think aloud?

Think aloud has been used to study search strategies and navigation behaviour of “people looking for detailed information”(Van Waes 1998). The author combines thinking aloud with an on-line camcorder recording mouse movements and browser action. This procedure is followed up by an interview, but he does not reflect on what the think aloud technique gives him access to as opposed to or integrated with other techniques! Think aloud is also used to understand mental processes in connection with writing programs (Bringham, John & Lewis 1991) and to uncover users’ performances and reasoning while engaged in learning Smalltalk (Koenemann-Belliveau et al. 1994). In addition, it has

been used to study students' writing and reading processes or the cognitive processes involved in problem solving.

Embedded in most studies is the understanding that the technique allows us access to the cognitive processes and mental behaviour, just as it gives us insight into thinking. Karsenty (2001) argues that thinking aloud procedures may be used to identify the cognitive processes responsible for users' behaviour. He points out that it cannot be used when users are involved in a spoken dialogue - in this case talking with a computer system over the phone. The author underlines the obvious, the technique puts a cognitive load on and requires a cognitive involvement from the user, which may interfere or even compete with the cognitive requirements of the interaction or the tasks. His interest is to identify the psychological sources of user variability, meaning "all cognitive factors affecting the assignment of meaning by users to the system's behaviour and utterances, and causing their reactions to the systems prompts" (p. 15). However, Katalin (2000) sees no problems with a double cognitive load. He used think aloud to get access to students' reading comprehension, "... the closest possible way to get to the cognitive processes of readers" (p. 1) and he understands text comprehension as a problem solving activity. He suggests using thinking aloud when students are engaged in reading in a second language: "Because it involves considerable efforts on the reader's part to make sense of a text written in an unfamiliar code" (p. 1). He does not consider it a problem that two different cognitive processes are running simultaneously, the process of reading and constructing meaning from a text written in a foreign language, and the process of having to talk and verbalise one's thoughts. Hence, he does not anticipate any cognitive interference in the process.

Yet, it has been argued by Jennifer Branch (2000) that thinking aloud or concurrent verbalisation are problematic when the task involves "a high cognitive load, when the information is difficult to verbalise because of its form..." (p. 379). Branch has compared the effectiveness of concurrent and retrospective verbal protocols in her study of adolescents searching information. She focused on participants' cognitive, behavioural and affective processes. Branch's report on the research is interesting as it initially discusses the application of verbal protocols in different fields and caution about its use. The paper is also exceptional because Branch gives a detailed description of the test setting, the facilitator's role, the instructions to the test-person and the steps in the test, besides detailing the analysis of the data, the steps and the procedure. She argues that think aloud provides "the most complete and detailed description of the information-seeking processes ...the specific search terms and decision-making steps...allowed a glimpse into the affective nature of the information-seeking process as well" (p. 382). Branch points out, however, "the reasons behind the decisions that were made were often explained in the Think After"

(p. 389), i.e. the interview that followed the think aloud session.

But what do we get access to when asking users to think aloud? Is it really an easy and straightforward technique? Does it really give us access to what goes on in people's heads? Boren and Ramey(2000) has questioned the technique and the theoretical underpinnings. They studied how practitioners actually carried out think aloud sessions and discuss the practice in relation to the classical work by Ericsson and Simon (1984). They argue that it is necessary to have a firm theoretical grounding and a unified practice before the technique can be called a method. They suggest exploring alternative theoretical positions and introduce speech communication as one possible approach, although there may be many others.

#### **THE CLASSIC: VERBAL REPORTS AS DATA**

Before proceeding any further, we want to introduce the readers to and discuss the work of Ericsson and Simon on verbal report as data. In the classic text from 1984: Protocol Analysis, Ericsson and Simon discussed the use of introspective data<sup>4</sup>. They wanted to reinstate verbal data as a valid resource for understanding human cognitive processes to make it (a) possible to use verbal data to verify, not only discover, phenomena of interest, provided (b) that verbal data was interpreted within a theoretical framework. They stated explicitly that they did NOT try to analyse peripheral thoughts, daydreaming, thoughts containing mental imagery and thoughts related to feelings.

Ericsson and Simon suggest that most performance measures rely on responses that are psychologically indistinguishable from a verbal report, as some kind of verbal reporting usually is necessary to understand people's actions, even in very simple tasks. They argue that a sentence is the verbal realisation of an idea, and verbs in a sentence can be used to identify different kinds of information and different cognitive processes. They distinguish between (classical) introspection, retrospective reports and communication to the experimenter on the one hand, and on the other verbalisation of currently "heeded" thoughts (thoughts reflecting current attention).<sup>5</sup> This differentiation is tied to

<sup>4</sup> Since then a revised version of the book has been published (1993), but the 1984 edition is the one that is cited widely.

<sup>5</sup> Under think aloud conditions, investigations have shown that 96% of verbalisations are concurrent. Raw data demonstrates that this heeded thought produce more pronouns and fragmented utterances, including more verbs in the present tense, while retrospective reports produce more verbs in past tense and introspection produces the speaker as the grammatical subject (I, my head), epistemic verbs (remember, feel, know) and lacks information on the current task (Ericsson and Simon 1984).

their understanding of short-term memory. The assumption is that everything we know has, at some point, gone through our short-term memory (STM) and we have been conscious of it. We can verbalise what we are perceiving while in the process of perceiving, and we can verbalise what we were conscious of if questioned shortly after the process has taken place. This is because it is still retained in our short-term memory. However, if there is a time span between perceiving and the request to recall, we will produce descriptions and explanations - not a report of our immediate thoughts, because the information from STM is lost.

As a consequence of this understanding – and because of the aim: to study task directed cognitive processes - only concurrent and certain kinds of retrospective verbalising will address the information employed while performing a given task. The authors' interest is to identify and analyse these verbalisations and in their model they distinguish between three kinds:

- Vocalisations of thoughts that are already encoded in the verbal form (talk aloud).
- Verbalisation of a sequence of thought that are held in memory in some other form, e.g. visually (think aloud).
- Other verbalisations (retrospective reports on thoughts not held in memory).

For Ericsson and Simon, it is primarily talk aloud and think aloud that are of interest, because these verbalisations express the content of short-term memory. Retrospective reports are only interesting as cross checks (not as anything that can make contributions), and only if carried out under conditions that may elicit information from STM.

According to the authors, talk aloud tasks call for cognitive processes operating directly on orally encoded information. The verbalisation begins immediately and proceeds as a vocalisation of internal speech. Protocols will take the form of a serial delivery of oral codes, e.g. numbers, letters, etc. The thinking aloud task calls for both orally encoded information and other kinds of thoughts held in short term memory. Protocol delivery will take the form of sentences that can be understood as thoughts without or within the context of other thoughts. Retrospective reports produce output similar to thinking aloud, except it is more coherent and more prone to errors if compared to what the subject actually saw and did during the session. For example “what am I thinking about” suggests that the subject only has infrequent access to intermediate stages in a thought process because of automation of the process or some kind of meta level or strategic thinking instead of “report of information attended to” (Ericsson & Simon 1984, p. 244).

### The protocol analysis

The main focus of Ericsson and Simon's work is really the formal analysis of verbal reports, and their method for

protocol analysis was developed with the purpose of modelling task directed cognitive processes. In their understanding, the analysis begins when a given theory has been used to identify the relevant part of the universe to be investigated, and the data has been collected in a raw form (audiotapes from the experimental session). The first part of the analysis involves writing and editing a transcript of the audiotapes, leaving out the information used to segment the verbal stream. The second and main part of the analysis is the actual encoding of the verbal reports. The encoding is done by using a priori determined coding categories. Each segment must be treated independently of the surrounding text. Hence, the method requires a high awareness of the relation between “raw data, interpreted data and theory” (p. 259). We will not go into details, but only briefly introduce the three steps.

The raw data is the first step, and Ericsson and Simon found that the emergence of new technology, the audiorecorder around 1945, “enhanced our ability to treat verbal protocols as hard data” (p. 4) - either in the form of audiotape or as verbal transcripts of the recorded tapes. The possibility of different researchers going back to the same verbalisations opened up for the development of more explicit theoretical assumptions based on the interpretation of data because: “data do not speak for themselves, especially in a system containing a memory that prevents observations from ever being exactly replicated. They must always be encoded and interpreted in the framework of a theoretical structure.” (p. 169).

The second step is the segmentation, where the authors seductively promise simplicity: “...verbal protocols are usually first segmented into individual statements (assertions, propositions). This simple encoding is seldom difficult or problematic” (p. 172). However, the segmentation is actually more complicated and in other parts of the book Ericsson and Simon go deeply into the methodological problems of segmenting verbal data: “...protocols must be divided up so that each segment will constitute one instance of a general process...”, suggesting that the appropriate cues may be: “...pauses, intonation, contours, etc., as well as syntactic markers for complete phrases and sentences – the cues for segmentation in ordinary discourse” (p. 205). Furthermore, they suggest that segmentation may be done on the basis of content, e.g. ideas or by time.

The final step is the actual encoding. An encoding scheme should be developed a priori and the vocabulary should be developed from (a) an initial task analysis and (b) “...from a preliminary examination of the protocols” (p. 266).<sup>6</sup> However, the authors note that some tasks have a

<sup>6</sup> As a good rule, Ericsson and Simon suggest that the number of terms in the encoding vocabulary necessary to encode 90% of the protocol will constitute five or ten per cent of the length of the protocol text. As long as it

precise language for communicating, while others do not share a common vocabulary. This is a recurring problem in many domains when trying to establish what goes on in people's heads. It is e.g. difficult to talk about cooperation in the maritime domain, because they use a command and control language, not a cooperation language.

#### Reflections on verbal reports

In retrospect, we notice that the wish of Ericsson and Simon to reinstate verbal data in scientific research (and this wish was in our opinion highly recommendable) induced them to develop a protocol of analysis in which all "noise" was discarded, leading to extremely rigid reductions. Their assumptions that only introspections, which are verbalisations of currently heeded thoughts, will enhance investigation into task directed cognitive processes lead to an information processing paradigm. The embedded understanding of human beings as verbalising, task oriented individuals, acting in splendid isolation with no context, no senses and no emotions to hold them implied losing the psychological being and substituting it for an information processing entity. Even in cognitive modelling, it has recently been suggested that we need more interactive and environmental modelling considering the different and complex bodily interaction with the task environment (Ritter & Young 2001). Besides, present day usability studies and research on human thoughts are not interested in task directed cognitive processes in Ericsson and Simon's sense, but rather in how thoughts are mediated by knowledge structures or artefacts that we design and use<sup>7</sup>.

One particular consequence of Ericsson and Simon's use of their model of verbalisation has led to recent criticism. It is their advice to instruct the subject to "keep talking" (Boren & Ramey 2000). Ericsson and Simon argue that in their model verbalisation will always lack behind thinking in time, except perhaps in the execution of very new tasks. This is because thinking in already encoded verbal forms is fast, the activation of "old thoughts" somewhat slower and only the generation of "new thoughts" is really slow<sup>8</sup>. That is why it is important to slow down the

thought processes (or speed up the verbalisation) and why they suggest: keep talking. The criticism addresses the evaluator-subject communication in Ericsson and Simon's model, but it seems to have overlooked that the closest Ericsson and Simon come to discuss how to interact with the participants is the "keep talking" advice. However, this is mostly dealt with in the appendix in their book. The true effort and focus of Ericsson and Simon were on how "raw data" is understood and presented in research. We might say that they were not interested in the technical and practical ways in which raw data is generated (see their recent work, Ericsson & Simon 1998), they were interested in the analysis of the protocols. The keep talking solution is explained as a means to circumvent the delay in time as verbalisations lack behind thinking. But how do we know that it is a question of speed and not a question of cognitive interference? Recalling that students reported that thinking aloud disturbed their interaction with the computers, we would venture to suggest that there is another possible explanation: two cognitive processes are competing, the process of thinking and the process of verbalising, hence the delay.

#### THE PSYCHOLOGICAL POSSIBILITY OF INTROSPECTION

The discussion of the value of introspection has a long history, and Ericsson and Simon's model should be understood in relation to an earlier paper on verbal reports by Nisbett and Wilson (1977). According to Ericsson and Simon, Nisbett and Wilson discredited introspection. But Nisbett and Wilson actually have the same goal as Ericsson and Simon: to distinguish illegitimate introspection from other kinds of more legitimate uses of verbal reports. While the two pairs of authors agree that "true introspection", in the sense of classical, *direct* introspective access to higher order cognitive processes, is impossible, they have different strategies for improving introspection. Ericsson and Simon focus on getting access to lower order cognitive processes like perception and memory, while Nisbett and Wilson focus on identifying the condition for giving accurate verbal reports on higher order cognitive processes such as "thinking", "affective appraisal" and "action systems". They see introspection as based on people's causal theories and judgements of plausibility.

Nisbett and Wilson conclude from a review study that people's reports on their cognitive processes are accurate reports, when influential stimuli are salient and plausible causes of the responses they produce. But the reports are inaccurate, when stimuli are not salient or plausible causes. According to Nisbett & Wilson, people cannot report accurately on their own cognitive processes (true introspection): "...we sometimes tell more than we can know...people sometimes make assertions about mental

---

(thoughts/minute) of 2-11, and (b) a rate of verbalisation (words/minute) of approximately 100.

---

is not important for the theoretical predictions, they argue that it is reasonable to use synonyms for terms in the task language.

<sup>7</sup> Ericsson and Simon point to this issue when they write about individual differences in ability to verbalize, e.g. gender. They cite a study "only highly verbal female subjects were able to give rich and fluent verbalisations in a reverie condition containing mostly visual images" (p. 250).

<sup>8</sup> Ericsson and Simon suggest that we should expect a verbalisation rate of hundred words per minute, close to a normal relaxed conversation, when the subject verbalises non-oral information. In thinking aloud (not talking aloud) one should expect (a) a rate of thinking

events to which they may have no access and these assertions may bear little resemblance to the actual events" (Nisbett & Wilson 1977, p. 247). But the errors that people make are systematic and regular and due to the application or generation of causal theories about connections in the situation. They suggest that if we want more accurate verbal reports, we should consider the psychology of the people we ask.<sup>9</sup>

According to Nisbett and Wilson, it makes no sense to talk about a risk of modifying the thought under investigation during verbal reporting, because the thought is the application of causal theories, the judgement. They see people as having or generating causal theories that may stem from different sources, which are mainly social and cultural: (a) explicit rules in a culture, like traffic regulations (b) implicit cultural rules, like dating rules (c) rules developed by an individual on basis of empirical observation (d) in situations where none of the other rules apply, e.g. an individual may develop new rules.

In their understanding, people will use such rules to generate explanations when they are asked to explain or describe what their thoughts were. Their explanation is exactly the kind of introspection that belongs to higher order cognitive processes, which the model of Ericsson and Simon exclude from the protocol analysis.

#### The practice of introspection

As indicated above, the analysis of protocols has been well discussed and the same applies to the theoretical possibility of introspection. However, there is a need to improve the theoretical underpinning of the interview part of the thinking aloud method. We will suggest that it pays to look closer at the "practice of introspection"

Vermersch (1999)<sup>10</sup> defines the practice of introspection as an individual's introspective activities governed by an expert interviewer, i.e. introspection in second person (what do you do...), instead of the traditional introspection in first person (I do...) or the verbal data approach that can be labelled introspection in third person (what does the subject do...). According to Vermersch

(1999), the practice of introspection includes the management of "relational dimensions, for genuine maintenance techniques ensuring an element of mediation in the course of the introspective process, which is however certainly present" (p. 11). The focus is on how to gain access to subjective experience, how to regulate the access and how to develop the competence needed to know that access of this kind has been gained. In this approach to introspection, the aim is to move phenomenology from the philosophical level to an established practice. In his discussion,

Vermersch never fully explains his view of human beings or his understanding of cognition. He does, however, focus on the adult human as a problem solver and how to gain access to the problem solving process. He states that access to subjective experience can be achieved – as in classical introspection - from the first person point of view, provided a stable state of attentive presence has been attained, which requires long training and hence excludes most subjects. Instead, Vermersch limits his interest to the practice of introspection, to what he calls "observation in the present" (as opposed to "presentification of past lived experience"), (p. 2). He points out that introspection is still with us, but under another name. Until 1940 it was called introspection, but after World War 2 a functional view of introspection gained ground and we now talk about introspection as verbalisation. The conditions for practicing introspection are: (a) there exists a disjunction between the logic of action and the logic of conceptualisation (b) it should not be necessary with exhaustive scientific knowledge to practice introspection, but it will require some practical experience with the method.

Vermersch criticises Ericsson and Simon for overlooking the need for the cognitive act that lies behind their "verbal encoding", "concomitant verbalisations" or "simultaneous verbalisation of the activity taking place", which supposedly should eliminate the "deformation, forgetting, and rationalisation" that a posteriori verbalisation brings with it. (p. 11). He argues that Ericsson and Simon are "concealing introspection under verbalisations alone" and that they dispense with the subject's point of view. To Vermersch, subjective experience is what counts: "introspection furnishes descriptive verbalisations in the second person of what can appear to the subject, within the limit of what has already been brought to conscious awareness or of what should be brought to conscious awareness" (p. 14).

He believes that his approach can avoid the "extremely impoverished subjectivity" that results from Ericsson and Simon's method. Vermersch argues that he – as practitioners before him – is able to "...take account of the sensorial modality in which a content of thought, a mental act, is experienced as being used" (p. 13). As Vermersch sees it, we need to consider introspection as an "act of introspection", a "reflective activity", which

<sup>9</sup> One way to get accurate reports, according to Nisbett and Wilson, is to allow people only to participate in session where the conditions for being accurate are present, that is: "...those for which the influential factors are plausible and are included in [the subject's] a priori causal theories" (p. 250)(1977).

<sup>10</sup> Pierre Vermersch is a cognitive psychologist who has made considerable research in operative intelligence theory and the understanding of adult intelligence in problem solving. He has been at CRNS in France since 1970. Vermersch has developed an original interviewing technique aimed at producing an a posteriori, introspective verbalisation of specific lived experiences.

includes considering "...the time of meta reflection, the description of the act, the originality of the reflective activity" (p. 18). When we introspect, there is an initial "...feeling of poverty, of indigence...", followed by a "...more or less durable vacuum..." and the subject who initially only has access to reflected activity feels that "...he knows nothing or at best a few banal generalities..." (p. 18). This "filling in" during reflective activity (introspection) takes place by stages and in accordance with a rhythm that is different from reflected activity (reasoning, thinking about something).

## DISCUSSION

To understand users and get access to what goes on in users' heads have proved difficult within the framework of the information processing model. Its conception of the user as a rational and goal oriented task driven entity, and its focus on analysis of "hard" data has done away with the living human subject. But the human being is not – not even when acting as a test user – an information processing entity. The human is a psychological being engaged in a psychological interaction, which cannot be reduced to that which is concurrently verbalised. Thinking is much more than what can be explicitly expressed in words.

To get access to human cognitive processes, a way forward may be to develop a practice of introspection; to expand our knowledge about the reflective activity of the user in the expert-guided think aloud session. One could also look for approaches that focus on the subject's culture, practice and activity. The cultural-historical position in Activity Theory, which seem to hold a unique position in the Scandinavian research tradition, offers such a model of cognition.

### Information processing and Activity theory

Activity Theory consists of a set of basic principles that constitute a general conceptual system rather than a highly predictive theory. The basic principles of Activity Theory include the hierarchical structure of activity, object-orientedness, internalisation / externalisation, tool mediation and development (Kaptelinin & Nardi 1997). The basic unit in Activity Theory is 'activity' (Bødker 1996; Kuutti 1996). Kuutti and Arvenon (1992) offer an enumeration of the properties that activities have with respect to the subject. They suggest that an activity has an active subject (actor) who understands the motive of the activity. In activity theory, this subject can be individual or collective and emotions form part of the activity. Emotions are interior signals and function as guide for actions and cognition (Flading 2002). Hence, activity theory obviously represents a possibility for a radical expansion of the information processing approach.

However, activity theory seems to have a tendency to exclude both emotions and the "mind" as a useful concept: "Activity Theory does not accept a dualistic conception of an isolated independent "mind"...a person's internal activity assimilates the experience of

humanity in the form in which it manifests itself in the corresponding external activity" (Kuutti 1996, p. 33). Although efforts have been made to emphasise the human actor as an autonomous agent (Bannon 1991), the subject's verbal expressions are seen as unproblematic and interpreted at face value from the perspective of activity theory (Turner, Turner & Horton 1999). Hence a question of tacit knowledge becomes irrelevant. Either we have an understanding of the user as an information processor or a treatment of the human consciousness as a by-product of goal directed tool use in work situations. In both cases, we have a limited view of humans, which does not consider many everyday forms of thinking, nor tacit insights, and disregards feelings and emotions.

### Experimental research and introspection

The reviewed literature has already pointed towards necessary enrichments of the think aloud sessions by using other techniques as well. However, in order to understand the scientific value of the techniques, explicit descriptions of the design and test procedure and the framework for analysis are required. Just as the thinking aloud technique itself needs to become the object of scientific investigations. One such is Boren and Ramey's study of how the technique is applied in practice. Another, we have already outlined: user testing and user evaluation of the technique. However, it may be worth looking closer at experimental psychology, and Ericsson and Simons model with the three levels of introspection: talk aloud, think aloud and retrospective thinking aloud may be an interesting starting point for enhancing our access to what goes on in people's minds.

With their Mindtape technique (Nielsen and Christiansen, 2000) have tried to take the study of introspections one step further. The authors studied computer supported collaborative interaction among a group of geographically dispersed geo-scientists. They did video recordings of distributed collaboration and their approach was to review a priori identified sequences of the tapes together with the users. Their approach may be seen as a way where introspection in third person (the analyst interpretations (or lack of) is combined with introspection in second person (the interviewer in the review sessions) and with introspection in first person (the user). The reviewing served as validation of the researcher's interpretations, but it was also dictated by their methodological approach which is to invite subjects to become participants in their own cognitive processes and not reduce them to research objects. Thus the main function of the review recordings was to act as mental trigger for introspection - in first person. The reviewing would make users recall, "... in extremely detailed grain, what they did, and why, what they expected to happen, what they thought when a visual image appeared on the screen, why they juxtapose another image etc. They seem capable of making internal thought processes explicit, and it is almost as if a detailed descriptions of their thought processes, a "Mindtape" of



their tacit inferences ....“ (Nielsen and Christiansen, 2000, p.2) was being replayed. They argue that such detailed verbalisations cannot be captured through a more traditional technique such as thinking aloud. Obviously a verbal protocol of a think aloud video recording, followed up with an in-depth interview may be able to produce a verbalisation of some of the tacit inferences, but users think faster than they can speak and their thoughts are much more complex than they can verbalize. Besides, the cognitive load of thinking aloud takes the users' focus away from the task they are suppose to do. Think aloud imposes constraints because the users constantly have to be aware of having to transform thoughts into words - the cognitive act of concurrent verbalisations. We may speak of a double cognitive load, and besides, many of the thoughts cannot be expressed in real time while the user is simultaneously interacting with the computer and the files, or taking part in a dialogue with a colleague. These thought processes take place much faster because they lie underneath or surround the language. (Kirkeby 1988).

The Mind tape technique seems promising because the processes of insight that runs while the user interacts with a computer, or collaborates with colleagues may become partly explicit. As it is not the users memory which serves as recall mechanism, but the recorded interaction with the interface, the actual sequence of events will structure the verbalisations, reducing problems with subjective recall. At the same time Mind tape seems to get around some of the problems with double cognitive load and cognitive interference because the introspection may invite both talk aloud, think aloud and retrospective verbalisations. The technique is still at prototype level, however it seems to invite introspection at all three levels and open for verbalisations such as retrospective reports (level 3) thinking aloud (level 2) and talk aloud (level1). Mindtape with its collaborative usability testing, seems to fall within the Scandinavian tradition, which is closely tied to collaborative design with developers and users, and a design process which is iterative and based on dialogue. Interestingly, however, experimental research on usability testing has not played a significant role here. Maybe because the Scandinavian HCI approach has taken the activity theory to heart and here the user's verbal expressions are seen as unproblematic and interpreted at face value. However, Scandinavian research may benefit from usability testing based in experimental psychology.

#### ACKNOWLEDGEMENT

The paper has been discussed in sessions in the HCI research group at the Department of Informatics. Thank you to all of you. Thanks are also due to the Danish Research Council for the research grant that enables the group to work with the ideas in the DIT (Design of Interactive Interfaces) project: ref. 9900403.

#### REFERENCES

- Bannon, L. (1991), From Human Factors to Human Actors: The Role of Psychology and Human-Computer Interaction Studies in Systems Design, in *Design at work: Cooperative Design of Computer Systems.*, eds. Greenbaum, J. & Kyng, M., Hillsdale: Lawrence Erlbaum Associates, pp. 25-44.
- Bedny, G. & Meister, D. (1997), *The Russian Theory of Activity: Current Applications to Design and Learning*, Lawrence Erlbaum, Mahwah, New Jersey.
- Boren, M. T. & Ramey, J. (2000), Thinking aloud: Reconciling theory and practice, *IEEE Transactions on Professional Communication*, vol. 43, no. 3, pp. 261-278.
- Branch, J. L. (2000), Investigating the information-seeking processes of adolescents: The value of using think alouds and think afters, *Library & Information Science Research*, vol. 22, no. 4, pp. 371-392.
- Bringham, B., John, R. & Lewis, C. (1991), Usability Testing of a Graphical Programming System: Things We Missed in a Programming Walkthrough, in *Human factors in computing systems conference proceedings on Reaching through technology*, New Orleans, Louisiana, United States, pp. 7-12.
- Buur, J. & Bagger, K. (1999), Replacing usability testing with user dialogue - How a Danish manufacturing company enhanced its product design process by supporting user participation, *Communications of the ACM*, vol. 42, no. 5, pp. 63-66.
- Bødker, S. (1996), Applying AT to Video Analysis: How to make sense of Video Data in HCI, in *Context and Consciousness*, ed. Nardi, B. A., MIT Press, pp. 147-174.
- Clemmensen, T. (2002), Interest in theory and method in a Danish community of usability specialists (in Danish). *Working paper, Copenhagen Business School, Department of Informatics, HCI research group.*, pp. 1-30.
- Dix, A., Finlay, J., Abowd, G. & Beale, R. (1997), *Human-Computer Interaction*, Prentice Hall.
- Duncker, Karl (1945) On problem-solving, in Dashiell, John F.: *Psychological Monographs, The American Psychological Association, Inc.*, Washington DC, vol. 58, pp. 1-114. (reprint, Westport, Conn. Greenwood, 1976).
- Ericsson, K. A. & Simon, H. A. (1984), *Protocol Analysis. Verbal reports as data*, Cambridge Massachusetts.
- Ericsson, K. A. & Simon, H. A. (1998), How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and

- explanations of thinking., *Mind, Culture, & Activity*, vol. 5, no. 3, pp. 178-186.
- Flading, Janne(2002) Usability expert at B&O – personal communication
- Hackos, J. T. & Redish, J. C. (1998), *User and Task Analysis for interface Design*, Wiley, USA.
- Kaptelinin, V. & Nardi, B. A. (1997), *Activity Theory: Basic Concepts and Applications*, Available: [<http://www.acm.org/sigchi/chi97/proceedings/tutorial/bn.htm>].
- Kari, K. & Arvonen, T. (1992), Identifying Potential CSCW Applications by Means of Activity Theory Concepts: A Case Example, in *CSCW 92*.
- Karsenty, L. (2001), Adapting verbal protocol methods to investigate speech systems use, *Applied Ergonomics*, vol. 32, no. 1, pp. 15-22.
- Katalin, E. (2000), "Please, keep talking": The 'think-aloud' method in second language reading research, *Novelty*, vol. 7, no. 3.
- Kensing, F. (1998), Prompted Reflections: A Technique for Understanding complex Work, *Interactions*, vol. January-February 1998, pp. 7-15.
- Kirkeby, O. F.(1988) Kroppens Tanke (The thought of the body), *Samfundslitteratur*, Copenhagen
- Koenemann-Belliveau, J., Carroll, J. M., Rosson, M. B. & Singley, M. K. (1994), Comparative usability evaluation: critical incidents and critical threads, in *Human factors in computing systems : "Celebrating interdependence"*, ACM Press, New York, NY, USA, Boston, Massachusetts, United States, pp. 245 - 251.
- Kuutti, K. (1996), AT as a Potential Framework for Human-Computer Interaction Research., in *Context and Consciousness.*, eds. Bødker, S. & Nardi, B. A., MIT Press, pp. 17-44.
- Mehlenbacher, B. (1993), Software usability: choosing appropriate methods for evaluating online systems and documentation, in *Proceedings of the 11th annual international conference on Systems documentation*, Waterloo, Ontario, Canada, pp. 209 - 222.
- Molich, R. (1994), *User friendly systems (in Danish)*, Teknisk forlag, København.
- Nielsen, Janni (1992), *Designers' Decision Making: Design of a User Interface for a Picture Search*
- Nielsen, Janni & Christiansen, N. (2000), Mindtape: A Tool for Reflection in Participatory Design, in *Participatory Design*, New York, pp. 303 - 313. *Database*, Esprit 3066, AMODEUS RP7/WP14.
- Nielsen, J. (1994), Estimating the Number of Subjects Needed for a Thinking Aloud Test, *International Journal of Human-Computer Studies*, vol. 41, no. 3, pp. 385-397.
- Nielsen, J. (1992). Evaluating the thinking aloud technique for use by computer scientists. In Hartson, H. R. and Hix, D. (Eds.): *Advances in Human*
- Processes, *Psychological Review*, vol. 84, no. 3, pp. 231-259.
- Preece, J. (1994), *Human-Computer Interaction*, Addison-Wesley, England.
- Preece, J., Rogers, Y. & Sharp, H. (2002), *Interaction Design: Beyond Human-Computer Interaction*, John Wiley & Sons.
- Ritter, F. E. & Young, R. M. (2001), Embodied models as simulated users: introduction to this special issue on using cognitive models to improve interface design, *International Journal of Human-Computer Studies*, vol. 55, no. 1, pp. 1-14.
- Rowley, D. E. (1994), Usability testing in the field: bringing the laboratory to the user., in *Human factors in computing systems : "Celebrating interdependence"*. ACM Press, New York, NY, USA, Boston, Massachusetts, United States, pp. 252 - 257.
- Simpson, M. (1990), How usability testing can aid the development of online documentation, *ACM SIGDOC Journal of Computer Documentation*, vol. 14, no. 4.
- Turner, P., Turner, S. & Horton, J. (1999), From description to requirements: an activity theoretic perspective, in *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, ACM Press, New York, NY, USA, Phoenix, Arizona, United States, pp. 286 - 295.
- Vermersch, P. (1999), Introspection as practice, in *The View from Within - First person approaches to the study of consciousness*, eds. Varela, F. & Shear, J., Imprint Academic, Exeter, UK, p. 320 pages. The chapter is also available from <http://cognet.mit.edu/MITECS/Entry/voneche.html>.
- Waes, L. v. (1998), Evaluating on-line and off-line searching behavior using thinking-aloud protocols to detect navigation barriers, in *Proceedings of the sixteenth annual international conference on Computer documentation*, ACM Press, New York, NY, USA, Quebec, Canada, pp. 180 - 183.
- Wright, P. C. & Monk, A. F. (1990), The use of think-aloud evaluation methods in design, *ACM SIGCHI Bulletin December 1990*, vol. 23, no. 1.
- Yeo, A. (1998), Cultural Effects in Usability Assessment, in *CHI 98, Doctoral Consortium*, pp. 74-76.