

Chapter 3

The think aloud method

3.1 Introduction

In this chapter we discuss the history of the think aloud method and the conditions under which the method can be effectively used. Some of these conditions play an important role in knowledge acquisition and will be discussed separately. This chapter ends with an overview of the analysis process which will be discussed in the following chapters.

3.2 History of the think aloud method

The think aloud method has its roots in psychological research. It was developed from the older introspection method. Introspection is based on the idea that one can observe events that take place in consciousness, more or less as one can observe events in the outside world. Some early psychologists, for example Titchener (1929), went as far as to claim that the events in consciousness were the actual object of psychology in contrast to the outside world which is the object of the natural sciences. In this view, psychologists study the type of events that take place in human consciousness and their causal structure just as other scientists study the events that occur in the outside world. The analogy between introspection and observation was taken quite far. A methodological principle was for example, that only well trained observers (experienced psychologists) were to be used as observers because they would be able to interpret the events in consciousness in the right way, just as a biologist who observes animal behaviour may notice things that an ordinary observer will miss.

Introspection has led to some successful research but there were also fundamental theoretical and methodological problems attached to it. The theoretical problems concern the model of introspection as perception of the contents of consciousness. This model makes a separation between the processes in consciousness and the introspection process itself, thereby suggesting that the latter is not accessible in consciousness. If that were true, then how is the introspection process accessed by the observer? On the other hand, if both are considered to be accessible in consciousness, a 'homunculus' problem is raised: is the introspection process itself subject to introspection? These questions could not be answered satisfactorily within the framework of introspection as perception of consciousness. As we discussed in Chapter 2, the solution that underlies the think aloud method is to assume a simpler process (*verbalization* instead of observation and interpretation) and to assume that only the contents of working memory are verbalized instead of the entire cognitive process.

A methodological problem with more severe practical consequences is that in the introspection view the research data are the events that take place in consciousness. These are to be analysed and explained. However, these data are fundamentally accessible only to a single observer, who also performs the thought process. This makes it impossible to replicate empirical studies and thereby to settle scientific discussions about thought processes. These discussions and the built-in limitation of the introspection method made psychologists turn away from the introspectionist method and associated theories. Because introspection was a central method in studying cognitive processes, this also meant that psychological research turned away from cognitive processes. This contributed to the rise of behaviourism in the 1930s. Behaviourism took the other extreme view. It banned all theorizing about processes that cannot be observed from the outside of the body, as speculation, with the exception of physiological processes. (Physiology was considered a related field.) The history of the introspection method in psychology has made psychologists suspicious of methods that resemble introspection. Note that we know now that this suspicion is not justified with respect to the think aloud method for two reasons:

- The think aloud method avoids interpretation by the subject and only assumes a very simple verbalization process.
- The think aloud method treats the verbal protocols, that are accessible to anyone, as data thus creating an objective method.

There has been a 'thin' line of research even in the 1930s and 1940s that continued to experiment with variations of introspective methods. The main

methodological advancement with respect to the introspective method was to treat the (verbal) reports as data, instead of the processes in consciousness. The advantage is that these data are open to inspection and interpretation for anyone. The theoretical model of the production of the verbal reports became less important, now that a working method became available. Interesting results were obtained with the think aloud method by for example Duncker (1945) and de Groot (1946 and 1965). Duncker analysed problem-solving processes in terms of memory search. He explained the sequence of possible solutions that people explore from an informal model of retrieval of relevant partial solutions from memory. These solutions were then accepted, modified or rejected by applying them to the problem and evaluating their implications. Without verbal data about the process this is clearly hard to investigate. De Groot was able to describe problem-solving by expert chess players as progressive refinement of a plan, using a large set of specific concepts and principles.

By the end of the 1960s the interest in internal cognitive processes grew very fast and thereby the interest in methods that can provide data about these processes. A major result was the work by Newell & Simon (1972), who used think aloud protocols in combination with computer models of problem-solving processes to build very detailed models. Using this methodology Newell and Simon were able to explain protocol data from a theory of human memory and assumptions about the knowledge that subjects could bring to bear on a task. This work had a major influence, because it showed that very detailed explanations of verbal data can be obtained. Although many psychologists were skeptical, the method gained more and more acceptance especially in the period from 1980 on, when computer simulations of cognitive processes became increasingly popular.

In the 1980s computer scientists began to develop expert systems. Using techniques from Artificial Intelligence, they demonstrated that it was possible to build programs that performed at an expert level of performance. Initially, however, there was no systematic way to obtain the knowledge from human experts. Knowledge engineers used more or less structured interviews with experts to obtain the initial version of a knowledge-base and in a later stage employed them as an oracle to repair errors in the program: a false solution that the program had found was shown to the expert, along with the trace of the solution process and the expert was asked to indicate where this had gone wrong and how the knowledge should be modified.

This method amounts to a combination of introspection and a structured form of prompting that suffers from the strengths and weaknesses of both: introspection is likely to miss many special tricks, heuristics, shortcuts and

special case solutions and structured prompting is suggestive and imposes a particular format on the knowledge. In the worst case, this may focus the expert on an aspect of the problem that is actually not relevant and it may distort the knowledge (if the format is inadequate). Some of the expertise is built up by experience and cannot easily be articulated and explained to an outsider. This led some researchers to use the think aloud method to elicit expert knowledge. If the expert can apply his knowledge in problem-solving then this becomes visible in the protocols. In this way it became possible to obtain knowledge of which the expert was not aware and the free (verbal) format avoids distortions and misrepresentations.

Currently the think aloud method is accepted as a useful method by a large part of the scientific community in psychology and it also has its place in the repertoire of many knowledge engineers. In section 2.6 we discussed factors that threaten the validity and completeness of verbal data. These factors were: invalidity due to disturbance of the cognitive process, invalidity and incompleteness due to memory errors and invalidity and incompleteness due to interpretation by the subject. What is the position of the think aloud method with regard to these factors:

Invalidity due to disturbance of the cognitive process: Does the additional task of thinking aloud change the cognitive process? Will a different process take place than without thinking aloud? Consider the following example. A psychologist is interested in the way in which people operate a power plant. He instructs the operators to think aloud while they monitor and operate the control panel of the plant. However, at some point an operator seems to forget her instruction to think aloud and quickly takes a series of actions. If the think aloud instruction is repeated and emphasized then the operator becomes a bit irritated and begins to act differently. It seems that she now performs the task differently. Finally emotional and motivational factors can result in a cognitive process that is different from the process that would take place during task performance without thinking aloud. There is not much evidence that thinking aloud adds much to the effect of being studied and evaluated that is inevitable in knowledge acquisition and experimental settings. In the next chapter we discuss how to minimize these effects.

Invalidity and incompleteness due to memory errors: Errors due to incomplete or false recall are essentially absent in case of the think aloud method. In any case they are not comparable with the errors that are caused by reconstruction processes in memory.

Invalidity due to interpretation by the subject: In psychological experiments no evidence was found that think aloud protocols are inaccurate in

the sense that people give incorrect information about the cognitive process concerned (other than occasional errors like those normally found in spoken language). This does not mean that protocols are easy to understand. As we already saw in Chapter 1, interpreting protocols can actually be quite difficult. As with other verbal reporting techniques, the form of the information and the verbal ability of the subject determine the quality of the reports. Sewing on a button or selecting spices for a sauce would not be easy to verbalize for most of us.

Although the think aloud method does not suffer from the threats to completeness and validity that play a role in the other techniques, it introduces two new threats to the validity of reports:

Incompleteness due to synchronization problems: Thinking aloud takes place concurrently with the cognitive process. A cognitive process takes longer when the think aloud method is used. This means that people are able to slow down the normal process to synchronize it with verbalization. However, subjects frequently report that sometimes verbalization does not keep up with the cognitive process and that their report is incomplete. This is consistent with the observation that occasionally protocols contain 'holes' of which it is almost necessary to assume that an intermediate thought occurred here.

Invalidity due to problems with working memory: If reasoning takes place in verbal form then verbalizing the contents of working memory is easy and uses no capacity of working memory. However, if the information is non-verbal and complicated then verbalization will not only cost time but also space in working memory because it becomes a cognitive process by itself. This will cause the report of the original process to be incomplete and it can sometimes even disrupt this process.

A related effect occurs when verbalizing the information in working memory is difficult and uses some of the capacity of working memory. For example, if the process operator wants to verbalize one of the items that quickly pass, she must keep it in memory while finding a suitable description of it. This memory space cannot be used for other information. This is a problem if she is used to quickly check several instruments and then consider the results. If she tries to follow her routine, then the capacity of working memory may be insufficient. Problems with working memory and synchronization can be recognized by complaints by the subject and interrupted verbalizations.

If you find that the think aloud method does not work well possibilities are to change the method, the task or the subjects. We have already discussed

the range of applicable methods. Below we discuss the selection of subjects and tasks.

3.3 Selecting subjects

3.3.1 Criteria for selecting subjects

Both subjects and tasks must thus be selected that the effect of possible disruptive effects of thinking aloud is minimized. The cognitive process in which we are interested should occur when the task is presented to the subject, disruption of the process by thinking aloud should be minimized and so should synchronization problems and working memory overload. Both in scientific research and in knowledge acquisition one does not always have a choice. Research may be directed at a particular kind of persons and we need a random sample of those because the results must be generalized over all persons of this kind. In knowledge acquisition it is often difficult to get access to an expert and one often cannot choose. Two important properties of subjects with regard to the applicability of the think aloud method are the degree of expertise and verbalization skills.

3.3.2 Experts as subjects

If the think aloud method is used for the elicitation of expert knowledge several problems are likely to occur. Expert knowledge is often partially 'compiled' in the sense that experts are able to perform a task very well, but that they cannot explain how they found the right answer ('I just saw that it had to be this'). The think aloud method makes some of this knowledge visible. On several occasions we observed that experts were able to make their knowledge explicit in a discussion afterwards about their think aloud protocols and our analysis of the protocols. However, as with regular subjects, experts that perform a task as a routine and very fast, are unable to verbalize their thoughts during this performance.

One has to take into account social and motivational aspects in psychological research with human experts. These factors may induce a person to behave more rationally in a psychological laboratory than in a more natural setting. Being supervised by a psychologist may influence the reasoning processes in experts when they think aloud. Experts can be secretive about their expertise and they may be reluctant to give someone else insight in their actual problem-solving behaviour. Most experts are well aware that they cannot

easily justify their answers (acting by their routine), but they may not want their audience to know this. Because of this, they may adopt a more rational reasoning style for the occasion of the knowledge acquisition session.

If you study expert behaviour with the goal of building a knowledge-based system, one could argue that the expert's rationalized problem-solving behaviour is not a problem, but a potential positive feature of a technique. Inspection or retrospection seem appropriate here because they yield more rational knowledge than the *ad hoc* reasoning that appears in practice. The target knowledge-based computer system should also be an optimal system rather than an imitation of the expert, including his weaknesses. However, in practice a 'rationalizing' expert may produce rationalizations that have no relation with the actual expertise and that are a much weaker basis for a knowledge-based system than the observed problem-solving behaviour. In particular, a rationalizing expert is likely to hide the unclear, poorly understood areas in his expertise, either in an effort to help the knowledge engineer by keeping things simple and avoiding the messy details, or by helping the knowledge engineer to find justifications for the answers. She may for example refuse to solve a problem because it has no justifiable solution even when she has a strong feeling about the best solution. It is better to use the think aloud sessions to acquire the expertise 'in action' and to find interpretations, explanations and generalizations quietly and systematically during analysis of the protocols. An expert should therefore be instructed that it is more important that the protocol is natural and direct than that it is comprehensible to the person taking the protocol (the knowledge engineer). Incomprehensible parts can always be cleared up afterwards, but missing thoughts and knowledge cannot be recovered. This means that one should take care to ensure the cooperativeness of the expert for the think aloud sessions. One way to achieve this, is by pointing out that this method may reveal patterns in his or her reasoning that are novel and by making clear that the expert will be involved in the analysis of the protocols.

3.3.3 Differences in verbalization skills

There are substantial differences in the ease with which people verbalize their thoughts. As we shall discuss in Chapter 4, a little training will help people to become more fluent, but differences remain, even after training. As a result some protocols will be more complete than others. If it is reasonable to assume that this skill is not associated with the cognitive process involved then one would of course prefer to select subjects with good verbalization skills. However, for most skills this is not known. In our experience, the quality of

verbalizations is not strongly associated with other properties that can easily be observed or measured. One possible exception is age. Young children usually find it difficult to think aloud. It is not clear if this is due to their verbalization skills, to the content of their thought processes or to the general difficulty of concentrating on a problem-solving task. Here too, the only practical approach is to try out the think aloud procedure in a pilot session.

3.4 Selecting problems

The considerations that we listed for selecting subjects also apply to the selection of problems for use in think aloud sessions. We already mentioned that certain tasks are less suited because they involve non-verbal information or because speed is inherent in the task. Other factors may also interfere with the think aloud task. For example tasks that involve verbal communication (for example, air traffic control, psychotherapy) are in their original form not suited for the think aloud method. Also within a task area that is suitable (for example, solving physics problems or architectural design) it is not easy to select a task which gives good data. Important considerations are:

- (a) Is the task at a level of difficulty that is appropriate for the subjects with respect to the cognitive process? The subjects should not be able to solve the problem in an automated manner. The task should be difficult enough.
- (b) Is the task representative with respect to the cognitive process involved? One is sometimes tempted to select an unusual problem because only that will be difficult enough. A risk is then that it introduces matters that are only marginally relevant to the cognitive process one wants to study. A practical way to overcome this is asking the assistance of another expert in selecting several problems that are both difficult and not too far fetched. After the experimental session one should also ask the subjects if they felt that the problems were unusual in any sense.

Because of time restrictions it is usually possible to apply the think aloud method to only a rather small set of problems. The best way to handle this is probably to combine the think aloud method with less time-consuming techniques that make it possible to get a picture of the generality of the results that were obtained with the think aloud method.

3.5 Summary

Not all tasks or subjects are equally suited for the think aloud method. The purpose of the method is to obtain data about a cognitive process. Therefore the situation should be such that this target process takes place in 'optimal form' and that it is 'verbalizable' in the sense that it involves verbalizable contents on working memory, does not proceed too fast to allow synchronization and does not cause working memory overload. These factors are relative to subject and task. The same task may be automated or verbalizable for one person but not for another.

3.6 Overview of the analysis of think aloud protocols

After discussing the verbalization process and methodological aspects of collecting think aloud protocols, we shall now turn to the analysis of the protocols. Here we give an overview of the analysis process as it will be discussed in the following chapters. Figure 3.1 presents the objects that play a role in the analysis of think aloud protocols:

Psychological theory of problem-solving: This is a theory about one or more aspects of human problem-solving. Think aloud protocols are relevant for theories about aspects of problem-solving that appear during the problem-solving process (and not only factors that influence it or characteristics of problem-solving performance) and that are accessible to verbalization (see Chapter 5).

Task analysis: Normative and competence models of problem-solving, describing the best way to perform a task and possible alternatives. If a task is complicated and a detailed model is required (for example, a computer program that can perform a task) then a task analysis itself is constructed following intermediate steps. For example, one may first construct an informal sketch of the model, next a more detailed design and finally a computer program (see Chapters 5 and 6).

Psychological model: A task-specific model of the problem-solving process that is the result of applying the psychological theory to the task analysis. The result is a model that predicts from the psychological theory and from the structure of the task (and the knowledge required to perform it) how people will behave when performing the task. Just like the task analysis, the psychological model can also be built in steps, from an informal model to a computer simulation (see Chapters 5 and 6).

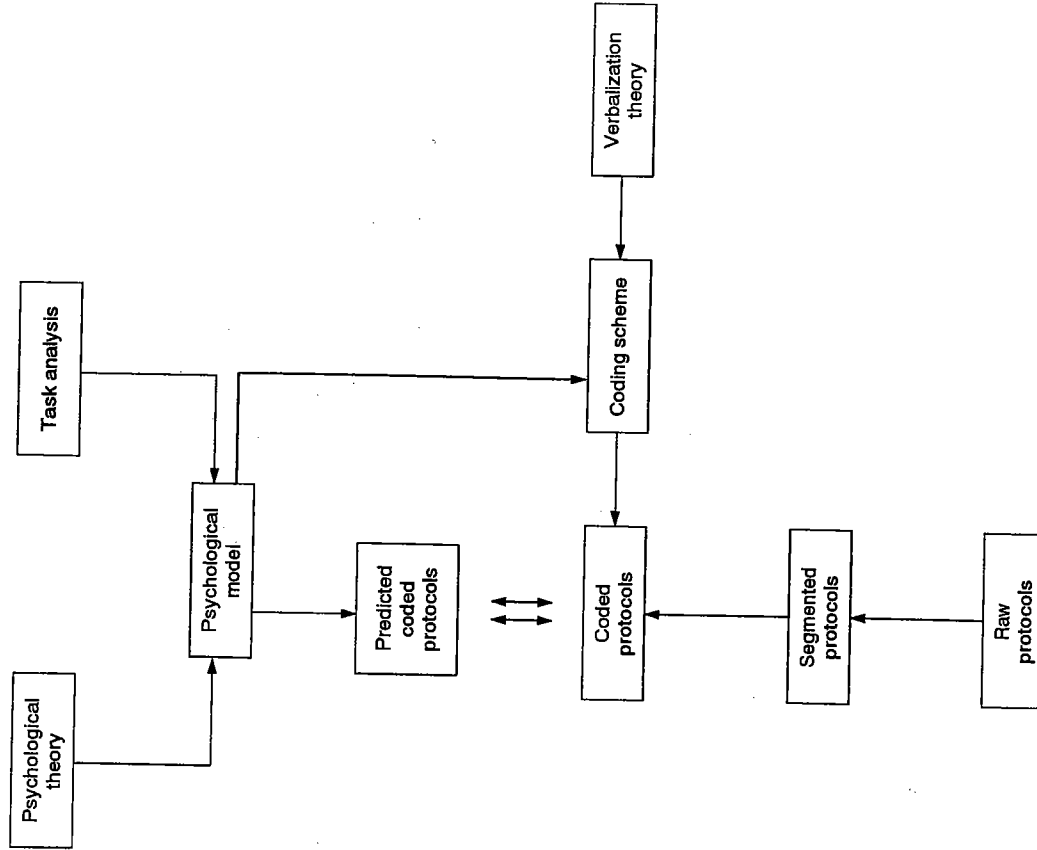


FIGURE 3.1: Overview of protocol analysis

Verbalization theory: This is a theory about the way in which thoughts that occur during problem-solving are verbalized. Verbalization itself is a task that has been studied in psychology and theories about cognitive processes can be applied here. This theory itself consists of a general psychological part (about the process of verbalization) and a part that is specific to the current task and the subject. Psychological theories about verbalization are usually not specific enough to construct a coding scheme and therefore it is necessary to use pilot protocols to obtain the vocabulary and phrasing that appear in protocols (see Chapter 7).

Coding scheme: An operationalization of the psychological model, that relates the psychological model to the text of the think aloud protocols. It is in the form of a coding scheme for protocol fragments. The coding scheme is obtained by applying verbalization theory to the psychological model (see Chapter 7).

Coded protocols: By applying the coding scheme to the protocols, coded protocols can be obtained.

Predicted coded protocols: The psychological model should imply predictions for the coded protocols (see Chapter 7).

Segmented protocols: The first step in the analysis is to divide the protocols into segments (see Chapter 7).

Raw protocols: This refers to protocols that are transcribed from the audio recording, possibly extended with other data such as notes or observations (see Chapter 4).

Literature

Pioneers of the use of the think aloud method in knowledge acquisition were Breuker & Wielinga (1987). The best source on the history and validity of the think aloud method is Ericsson & Simon, 1993. This gives an extensive discussion of the psychological literature on the method.